CAPCO

a wipro company

THE CAPCO INSTITUTE
JOURNAL
OF FINANCIAL TRANSFORMATION

# ARTIFICIAL INTELLIGENCE

#58 NOVEMBER 2023

# THE CAPCO INSTITUTE

## JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

# CONTENTS

## TECHNOLOGICAL

# OPERATIONAL



# ORGANIZATIONAL

# DEAR READER,

As the financial services industry continues to embrace transformation, advanced artificial intelligence models are already being utilized to drive superior customer experience, provide high-speed data analysis that generates meaningful insights, and to improve efficiency and cost-effectiveness.

Generative AI has made a significant early impact on the financial sector, and there is much more to come. The highly regulated nature of our industry, and the importance of data management mean that the huge potential of AI must be harnessed effectively – and safely. Solutions will need to address existing pain points – from knowledge management to software development and regulatory compliance – while also ensuring institutions can experiment and learn from GenAI.

This edition of the Capco Journal of Financial Transformation examines practical applications of AI across our industry, including banking and fintechs, asset management, investment advice, credit rating, software development and financial ecosystems. Contributions to this edition come from engineers, researchers, scientists, and business executives working at the leading edge of AI, as well as the subject matter experts here at Capco, who are developing innovative AI-powered solutions for our clients.

To realize the full benefits of artificial intelligence, business leaders need to have a robust AI governance model in place, that meets the needs of their organizations while mitigating the risks of new technology to trust, accuracy, fairness, inclusivity, and intellectual property. A new generation of software developers who place AI at the heart of their approach is also emerging. Both GenAI governance and these 'Developers 3.0' are examined in this edition.

This year Capco is celebrating its 25th anniversary, and our mission remains as clear today as a quarter century ago: to simplify complexity for our clients, leveraging disruptive thinking to deliver lasting change for our clients and their customers. By showcasing the very best industry expertise, independent thinking and strategic insight, our Journal is our commitment to bold transformation and looking beyond the status quo. I hope you find the latest edition to be timely and informative.

Thank you to all our contributors and readers.

Lance Levy, **Capco CEO**

# TECHNOLOGICAL

# OVERVIEW OF ARTIFICIAL INTELLIGENCE DEPLOYMENT OPTIONS

**ALI HIRSA** | Professor of Professional Practice, Department of Industrial Engineering and Operations Research, Columbia University, and Chief Scientific Officer, ASK2.AI

**SATYAN MALHOTRA** | Chief Executive Officer, ASK2.AI

## ABSTRACT

Artificial intelligence is a very powerful application whose time has come. At a quick glance, it can be really seductive to believe, for example, the purveyors of xxxGPT, that its deployment is as simple as pushing a button or is a "data in, miracles out" strategy. However, harnessing it effectively requires navigating a myriad of options embedded within its critical pillars of data, models, and visuals. The complexity is accentuated by the deployer's capabilities and the organization's openness to change, as outcomes move from rules to an objective-based spectrum. In navigating these challenges lies the key to optimal deployment.

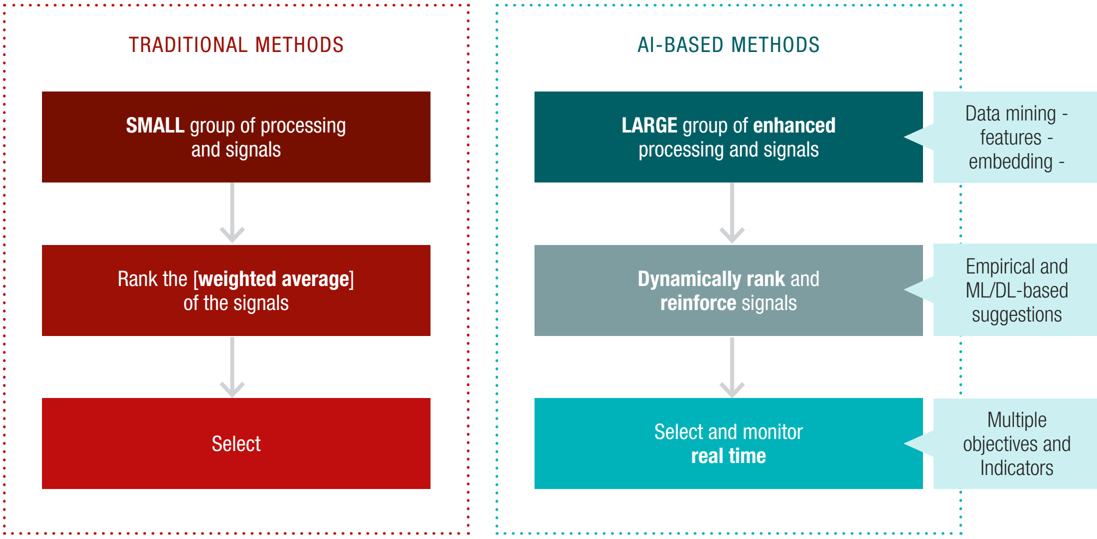
## 1. INTRODUCTION

In the realm of artificial intelligence (AI), some amazing things are being done by some amazing people, which is leading to some amazing results. Hopefully, this pacifies the shallow learning experts. Now for the deep(er) learning aspects of the current push of **AI everywhere and for all**. For the older engineers, most of the models being deployed (with some updates) have been there for many years, so what gives? Well, for one, we know that great strides in readily available computing power have been an excellent catalyst for the more ubiquitous push of AI. Another has been the ever-increasing money supply via the venture community and their ability to sell assets between themselves or to the next tier private or public community. This race to automate all things human is making engineering cool again with more jobs available than degrees being printed. Therein, establishing the current cycle by pointing the research engines at the opportunities at hand.

For some of us it has a hint of the Y2K[1] days when every boardroom, chatroom, money room, and classroom was pushing for solutions to the elusive double zero so the world would not come to a standstill. Since the world needs the engines running, it led to great billable hours for participants, spawning many startups, mixed solutions from developers, and money thrown at buzzwords. Pressure from time and money deployment compressed the processes for separating the grain from the chaff, which led to some not-so-pretty conclusions when the meals were finally consumed. History repeats itself, albeit maybe with some twists, tweaks, and turns. Similar compressed and expedited processes were seen during other such hyped times, including the dot com bubble. On the AI front, on February 9th, 2023, Google lost U.S.$100 billion dollars of its market value as the market punished it for its tardy Bard presentation. But was it tardy or is there an "AI catch up or lose" issue being exemplified? We found it interesting that some venture capitalists began pointing to the lack of practicing/checking the pitch before the presentation. Cannot see the forest for the trees or calculated censure given the exit plan relationship? The point remains that the pace with which all things AI are being pushed to consumers, users, funders, advisors, et al. seems to rekindle some memories for those of us who were at the table during the past frenzies. This time again, perhaps, maybe a little bit of overuse, abuse, and misuse of AI and its applications?

[1] For the younger readers, Y2K was the year 2000, when all computing was to come to an abrupt end due to the perceived poor programming of the elder generation. A lot of money was spent trying to fix the issue and we will never know if there were more issues or billable hours.

**Figure 1:** Approach comparison

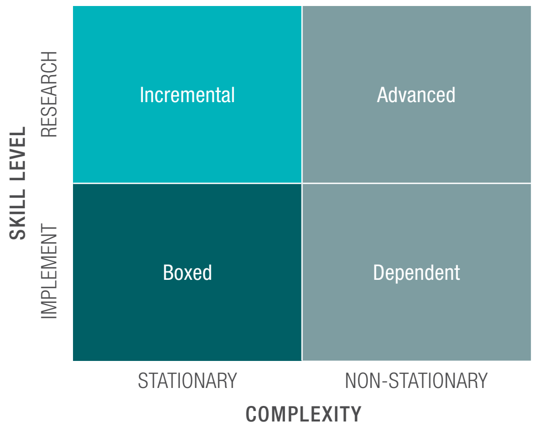



We are not saying that AI is not here to stay and note that the "rocket" train has left the station. Figure 1 visualizes the traditional versus AI-based deployments that allow parsing through numerous datasets, models, viewpoints, and visuals, etc., concurrently to continually assess historic and/or predicted performance within multiple aspects, subject to defined or suggested objective and evaluation functions.

The approach is very powerful and at a quick glance it can be really seductive to believe, for example, the purveyors of xxxGPT, that AI deployment is as simple as pushing a button or is a "data in, miracles out" strategy. Maybe true in the elusive future, but the current reality is that AI deployment has a lot of optionality and right choices need to be made to capture its immense potential. Being on the wrong frenzy driven side may entail wasted time and effort, as will be realized by some blindly following herds. The needs and level of intensity varies across use cases, regions, industries, etc. One way to think about this is in Figure 2, where for effective AI deployment we need to balance the expectations of the use case depending on the skill/experience level and the potential task complexity.

Within each of the quadrants, there is a lot of optionality and need for a lot of decisions. For the more complex cases, effective AI deployment is even more difficult and a lot of work still remains to be done. However, once any breakthrough happens then the trajectory of AI deployment engineering is relatively rapid, but we need those nudges and breakthroughs.

**Figure 2:** AI effort quadrants




For example, in the financial markets, given the current AI environment, one could consider using Natural Language Processing (NLP) techniques to read and file documents as somewhat of a boxed case, whereas, implementing NLP techniques to read real time data for advantageous financial trading so far remains an advanced case. Another way to look at this is that applying boxed cases in the non-stationary/research quadrant may be punitive and using advanced cases in the stationary/implement environment may not be value add. We will not get into the mathematical aspects, but where applicable, we will highlight limitations of assessed AI techniques as well as our proposed research nudges.[2]

[2] https://www.ask2.ai/research/

## 2. UNDERSTANDING OPTIONALITY

AI deployment is in a transitory stage and effective rollouts will depend on the knowledge bank of the deployers and decisions made by the users. As with Y2K, or the dot com bubble, or "then some", participants may burn money as AI is the new big idea whose time has come. Our suggestion is that if you have a seat at the table, it may be wise to look at the ingredients, chefs, and the dishes more closely. The more complicated dishes may need more than a naïve attempt from a cookbook and an independent mechanism for judgment. To make this discussion more pointed, we explore the AI deployment optionality around selecting a mutual fund manager. We assume those reading have some exposure to mutual funds and would ask you to parallel your current assessment processes. We look at the deployment optionality within each of AI's three pillars: data, models, and visuals.

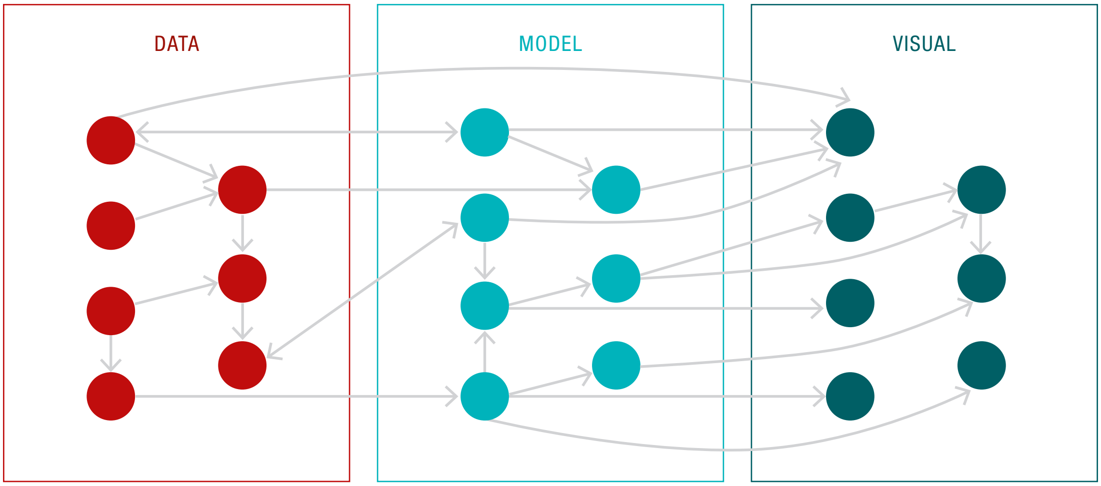
We stay away from the arguments around why these three or which pillar is more important. Some consider data to be the new oil, some modeling secrets to be the sauce, and some unique visual wrappers to be the trust builders. Additionally, judgment mechanisms have to be set that help evaluate the optionality within stability frameworks that allow evaluating the tracking error among other elements. Since this is done over time and on out-of-sample data, the judge is considered an independent unbiased framework for evaluating the results. As we will see, there is optionality there too as decisions need to be made on setting the appropriate objectives and associated evaluation criteria. This can be complicated, as judgments involve decisions on the related value system that supports the recommendations, selection, rewards, and penalties. The use cases, participants, judges, etc., are different and thus the deployment optionality needs to be understood for appropriate selection. This is because beyond considering all as vital pillars, each use case could have a very different path to "its" optimal solution. Finding that path or the tuning is possibly the key. In Figure 3, we had some fun illustrating some possible ways of connecting the dots.

Not surprisingly, there are numerous ways to connect the dots. Arguably, for boxed cases they may be established (or more or less specified for a use case), but until the advanced cases become boxed, the paths have to be tuned or are open to arbitrage. The arbitrage comes from the choice of faster deployment of any model/system (e.g., maybe untested for the use case) or cautious deployment of better models (e.g., tuned for the case). Depending on the use case the risk-rewards can be very different.

Furthermore, for binding the pillars, another layer comes into play, such as change management. As we know, without having the right people, entities, aspects, etc., be part of the AI deployment journey, or the right setup, there is a risk that all of the work may end up gathering dust somewhere. This is especially important as you begin to appreciate the level of optimality and decision points that needs to be addressed across AI deployment. Then, there are other related challenges of regulatory frameworks, local challenges, accessibility, computing power, etc., and many of them are transitionary.

Overall, for maximum effectiveness, it may be best to try not to compress the processes for separating the grain from the chaff. So, let us see what the cooking optionality entails.

**Figure 3:** Path choices



## 3. DATA

Data: the ingredients for the dish is a critical pillar. Simply put, although the xxxGPT purveyors may give an illusion, as per our assessment, models are not clairvoyants (yet). From a data perspective, it is as essential to have the data as well as to understand how to use the data. The current stage of AI deployment has put more focus on the "having part" with arguments spanning more is better, unique is better, etc. This is not unusual, as we are somewhat in the early stages of AI deployment and thus resource gathering is vital. However, the resources are not as readily, cleanly, or widely available. This opportunity has spawned providers from the software participants pitching streamlined storage or access capabilities, better processing capabilities, pure data providers pitching clean(er) or proprietary datasets, hybrids pitching better signal processing, and so on.

Until we reach the elusive utopian data stage, we find that data management straddles all the boxes of the AI effort quadrants. This becomes especially true as processes evolve from we "have" data to "how best to use" the data stage. Even within the "we have" data part, you will note that you need to be careful and know what you have/get as not everything is as simple as a pitched boxed case.

### 3.1 Nature of data

Data itself can be classified as "structured" or "unstructured". Structured data is tailored and generally stored in designated formats, while unstructured data is an amalgamation of different data types stored in their native formats. For example, your hard drive may be structured, but we can assume that various types of files are stored in the sub directories. This analogy can be extended to the task in hand and, as you will start to note below, the ability to manage both data types is generally a value add. Definitionally, processing unstructured data is more of a challenge and the key resides in effective and accurate extraction. Not surprisingly, a lot of effort is being expended in streamlining unstructured data so more and more can be part of boxed cases.

### 3.2 Types of data

The types of data sourced are topical. For our mutual fund example, we can source processed or unprocessed price data, holdings data, alternative data, news data, social data, proprietary data, and so on. Within each of these data types, there are various fields with varying frequency that all add to the data management complexity (e.g., multidimensional information can be tick level, minute, hourly, monthly, semiannual, etc.).

### 3.3 Storage of data

The traditional usage and familiarity are generally around "relational" databases, where tables are used to store data (think Excel). As relationships become more complex, "graph" databases may be better suited (think trees, branches, and leaves). Each branch or leaf can store various types of information, and since the types are somewhat grouped (e.g., within the branch or leaf) the number of connections is reduced versus a relational table where data is in a tabular form. For example, this can improve the response and management time associated with the queries. For our mutual fund example, the price and related information may be in a tabular form in the relational database, whereas connected information, such as alternative or social information, is in the graph database.

Data volume also has to be balanced with concerns around control and security, where fragmented data is harder to protect consistently. For example, the large data needs of LLM/GPT are understood but it is undecided whether to store data internally or use open-source solutions. The E.U.'s GDPR dictates data privacy norms and this puts an increased burden on data walls and mirrors, navigating global versus local datasets, inherent biases, etc. Basel regulatory pressures include making data auditable and reproducible for third parties. And so on.

### 3.4 Pre-processing data

The objective here is to have the data ready for analysis. The data can be sourced from a single or multiple sources, be in different formats, have different information, stored in a variety of ways and so on. For our mutual fund example, we also have to deal with multidimensionality and time series that are continually updated. To get the data ready for analysis, varying degrees of pre-processing may be required.

#### 3.4.1 APPENDING AND CHECKING DATA

Most participants take the source data as a given. Unfortunately, there is usually no one true source of data. For the mutual fund example, data can be received from multiple sources, can have different identifier codes/symbols (e.g., CUSIP, ISIN, SEDOL, TICKER, etc.), and could be subject to very different taxonomies and protocols. It is imperative to know what you are working with, the rationale for the difference, where the pitfalls are, and so on. For example, a comparison of data for the same mutual fund from two reputable sources can show different (a) alternative data – expense ratio compositions, b) price data – total return on how they capture capital gains and dividend days, or c) holdings data – sector exposures. It is worth noting that this is for financial products, where the reporting is more or less structured – as in regulation heavy, legalized via prospectuses, and reported via electronic exchanges. Now imagine these pitfalls where the data is unstructured and all of it could be driving exposure, sentiments, signals, and so on.

#### 3.4.2 FORMATTING AND CLEANING DATA

As the baseline data needs are set, additional steps are format cleaning and data standardization. Format cleaning requires streamlining the data, where some features may be stored as strings, could be nested, not have values, different frequencies, and so on. Once general formats are set, the dataset could further require "imputation" (e.g., filling the missing data with substituted values, where imputing time-series data should avoid data-leakage), NLP (e.g., aligning nomenclature that points to the same), and model specific engineering (e.g., standardizations). For example, filling missing data utilizes techniques from the simplest to the most sophisticated, including (a) forward autofill, (b) linear interpolation, (c) cubic spline interpolation, (d) cubic B-spline interpolation, (e) Brownian bridge, (f) variance gamma bridge, (g) Fourier transform techniques, and so on.

For the mutual fund example, the data is received at discrete points in time. We have to keep track of manager history across accounts rather than just continuity in the fund. When a manager leaves or joins another fund, the system has to account for the adjustment in expertise. Similarly for illiquid assets, the performance is self-reported as there are no central clearing systems. The challenge extends to managing incomplete data, incorrect data, reported biases, and so on.

#### 3.4.3 TESTING DATA

We have to ensure that the datasets are robust enough to deal with AI models, e.g., raw or processed with cleaned values. Testing includes ensuring perturbations, different signal-to-noise ratios, adversarial attacks, and such, do not drastically distort results.

### 3.5 Processing data

The assumption at this point is that the data is clean and readily accessible for analysis. The objective of this part is then to make the data ready for modeling purposes.

#### 3.5.1 DEALING WITH LARGE DATASETS

Generally, models are able to deal with more data better than with less data. At the same time, feeding similar data would lead to overfitting, auto-correlation, and other not so pleasant issues. Dimensionality reduction, such as "principal component analysis" (PCA), is one such method that can transform and reduce the number of measures or times so a single series can represent a set (without losing any information). However, if the datasets have time and multidimensionality aspects, then the standard PCA techniques may not give stable results. Here, we propose the robust rolling PCA (R2-PCA) that mitigates commonly found obstacles, including eigenvector sign flipping, temporal instability, and managing multiple dataset dimensions. If the objective is to identify some latent relationships or interrelationships among variables, then "factor analysis" (FA) can be the preferred method.

#### 3.5.2 DEALING WITH SPARSE DATASETS

This can be a real challenge for AI models, as nothing can be done without data. However, if there is some level of data, then that can be augmented with synthetic data. Techniques such as Bayesian sampling and adversarial generative modeling can help create data that closely mimics existing datasets. Here, generative and hierarchical models are used to sustain statistical properties and stylized facts for different frequencies in both the time domain and frequency domain. These are high-risk areas, as care must be taken to ensure that the augmented datasets do not contribute to an alternative reality. We refer to our work on "temporal attention" and "temporal transformer generative adversarial networks" (TAGAN & TTGAN), where images inspired the original work and now the work is being extended to account for various datasets including financial products.

#### 3.5.3 ASSESSING DATASETS

These techniques can augment the analysis and make for easier explainability when reduction or performance assessment techniques are applied across categories versus across the whole dataset.

- **Categorizing/classifying:** this is a simple form of grouping datasets where items can be bunched within predefined categories. This can be done using basic definitions or some manual structure of commonalities. In our mutual fund example, the industry has grouped funds using predefined classes, e.g., Large Cap, Small Cap, Fixed Income, and so on.
- **Segmenting:** this is a way to divide the data into parts or segments based on motivation. In our mutual fund example, it can be those funds that perform well during a regime.

- **Clustering:** this is a more advanced form of categorization/classification, where the groupings are based on similarities and data characteristics. For example, clustering mutual funds based on holding data exposure (equity, sector, etc.), performance measures, factor sensitivity, macro/market conditions, alternative data, etc. This can be done in a parallel or in a sequential manner. If datasets have time and multidimensionality aspects, standard "clustering" methods (K-Means, Hierarchical, etc.) may not give stable results. We propose CPK-Means and SIK-Means methods for producing stable and deterministic clusters over time.
- **Regimes:** this classifies the dataset into periods of similar behavior or events. The classification includes the defining characteristics of the regimes as well as the transition probabilities as movements between regimes. This can effectively assess the anticipated behaviors at similar points in time. For example, an advanced use case assesses how the fund clusters behave within and/or across regimes. Given the complexity of managing times series and multidimensionality, we propose an AI-based methodology for classifying regimes that produce stable financial regimes with transition probabilities.
- **Measuring performance:** these are constructed by manipulation of the same underlier in the form of a time series. For our mutual fund manager selection, the underlier is price, and the performance measures range from primitive to those requiring advanced financial engineering. A survey and a taxonomy of portfolio performance measures reveals that there are over a hundred such performance measures, and there is an assessment choice of time horizons (e.g., monthly, quarterly, annual, three years, five years, etc.) and roll/look back windows (e.g., daily, monthly, quarterly, etc.). Some of these performance measures are relative and thus need a designated benchmark. To quantify the manager's risk attitude, we propose an additional golf inspired "advisor assessment framework" with a scorecard, fairway average, and handicap.
- **Indexing/labeling:** this is a way of naming the grouped data to be easily referable. In our mutual fund example, Cluster A can be funds with high returns and Cluster B can be funds with low volatility. Note that since the data within the group takes on the implied meaning, this can lead to biased results and potentially amplify issues.

> "*Great experiences blend the exterior with calibrated power under the hood.*"

Given the inherent probabilistic nature of AI models for making the suggestions, recommendations, selections, and so on, it should not be surprising that a large number of the data processing techniques are statistical in nature. The key is choosing the proper technique and understanding that many of the boxed solutions may not work for the learning-based models.

## 4. MODELS

Models: the equipment for the dish are a critical pillar. They encapsulate the analytical part of the task and objective. This is a complex part of the deployment process, yet a lot is taken for granted or assumed to work, potentially as black boxes. Depending on the task, model dependency or deployment can easily straddle all the boxes of the AI effort quadrants. Attempting to naively transplant models across use cases with differing nuances, datasets, temporal considerations, dimensionality, and so on, can be punitive (depending on the appetite for the error rate). There are many models to choose from and some are better for the task, some easier to comprehend, some easier to explain, some easier to implement, and some less computationally expensive, etc. We need to be able to choose the "right" models and have mechanisms to know when they are working and when they are not working.

### 4.1 Setting the framework

The model deployment framework consists of setting the objective, measuring the results, evaluating the results, accepting or rejecting the results (or the penalty-reward functions for the more advanced models), and refining the models – then repeating the loop. In actuality, this is done rapidly and concurrently by running numerous models under various scenarios, parameters, assumptions, targets, etc. All are obviously subject to the deployer's knowledge, data depth, and available computational power.

#### 4.1.1 OBJECTIVE FUNCTION

The objective acts as the desired result for the model. This target cannot be abstract and has to be set as a quantifiable objective function. For example, if the results of the target are to drive a decision, then the target can be one or many steps towards "suggesting" the optimal decision. Consequently, the objectives can be interim or final, near or longer term, whole or components, sequential or nonsequential, single or multiple, and so on. Additionally, a subtle difference between an objective and a constraint is worth noting, where constraints are guard rails that drive the model towards the target.

#### 4.1.2 MEASURING RESULTS

This entails assessing if the objective or the target has been met. The selection, ranking, or recommendation is usually based on the results closest to the target. The more precisely the objective function is defined, the easier it is to measure the results. It is also essential to assess details around the results, for example, which models were performed, what the error rate was, under what circumstances or scenarios it was met, what features drove them, were there any outliers, and so on, as all this comes into play via the refinement loop.

#### 4.1.3 EVALUATING RESULTS

This entails accepting or rejecting the results. If this includes potentially rewarding or penalizing the results, then it also allows for setting the degrees of reward-penalty functions.

- **Back-testing:** this technique involves splitting the dataset into a training set and a test set. The model is trained on the training set and then evaluated on the unseen test set to assess its generalization performance. This evaluation helps determine how well the model performs on new, unseen data. This can also include various types of "scenario analysis", "stress testing", and "simulations".
- **Validation set and early stopping:** in cases where models have hyperparameters to be tuned, a validation set is often used. It is separate from the training and test sets and is used to evaluate different hyperparameter configurations. Early stopping is a technique that monitors the model's performance on the validation set and stops training when performance deteriorates, thus preventing overfitting.
- **Robustness testing:** this involves testing the trained model over various different data with known and similar characteristics to see how the trained model behaves. This could include having different degrees of noise, perturbations, and adversarial attacks.
- **Deployment:** once an AI model is deployed in real time applications, ongoing monitoring and evaluation are essential. This involves tracking the model's performance, detecting anomalies or drifts, and ensuring it continues to perform as expected. This also serves as input for the refinement loop.

### 4.2 Selecting the models

There are many types of models, including simplistic ones, complex ones, and those that auto choose between models. One more characteristic has to do with the representation of data or input, where if the data is presented in multiple levels and a different model is used at each level and gets combined for final decision making, then the models are hierarchical.

#### 4.2.1 RULE-BASED MODELS

These are the simplest form of models that operate based on predefined rules. They follow "if-then" logic to make decisions that are essentially fixed equations to represent relationships between inputs and outputs. These models are straightforward to implement and suitable for simple problems but are less effective for complex tasks and have limited flexibility and adaptability. A mutual fund selection example would be to select a fund if the total return is more than a certain percentage.

#### 4.2.2 REGRESSION-BASED MODELS

These models are suitable where there is a need to identify some form of a relationship between the inputs and outputs.

- **Linear regression:** these utilize a linear equation to model the relationship between input features and the target variable.
- **Lasso regression:** these perform feature selection and regularization by adding an L1 penalty term (the sum of the absolute values) to the loss function.
- **Ridge regression:** this incorporates an L2 penalty term (the square root of the sum of the squared values) into the loss function, encouraging smaller coefficient values.
- **Elastic net:** these combine L1 and L2 penalties, offering a balance between feature selection and regularization.

Regression-based models offer advantages such as interpretability and flexibility and allow for the assessment of the relative importance of different input features in determining the outcome. Regression models provide a statistical framework for inference and hypothesis testing and explicitly define assumptions about the relationships between variables, which helps guide the modeling process. Regularization techniques, such as ridge regression or lasso regression, can be applied to mitigate issues like multicollinearity or overfitting. They are computationally inexpensive and provide a baseline for comparing the performance of other complex models. They are also widely understood and used in various fields, making them accessible to researchers and practitioners, e.g., credit scoring, demand forecasting, econometrics, marketing analytics, risk assessment, and in general predictive analytics. A mutual fund selection example would be to select a fund based on the regression coefficients of measures, where it can be expected that the coefficients would adjust for the changing performance.

#### 4.2.3 BAYESIAN MODELS

These, also known as belief networks or probabilistic graphical models, are graphical representations of probabilistic relationships between variables. They use directed acyclic graphs to depict dependencies and conditional probabilities. Bayesian networks are used for reasoning under uncertainty, probabilistic inference, and decision making.

- **Bayesian networks:** these extend traditional linear regression by incorporating prior distributions over the regression coefficients. It provides a probabilistic framework to estimate the uncertainty associated with the regression parameters and make predictions.
- **Bayesian linear regression:** these extend traditional linear regression by incorporating prior distributions over the regression coefficients. It provides a probabilistic framework to estimate the uncertainty associated with the regression parameters and make predictions.
- **Gaussian processes:** these are flexible probabilistic models that define a distribution over functions. They can be used for regression, classification, and uncertainty estimation. Gaussian processes capture prior assumptions about the smoothness and correlations in the data.
- **Variational autoencoders (VAEs):** these combine the concepts of autoencoders and Bayesian inference. They use deep neural networks to learn a low-dimensional representation of the data and model the underlying distribution in a probabilistic manner. VAEs enable the generation of new samples and provide uncertainty estimates.
- **Bayesian neural networks (BNNs):** these integrate Bayesian inference with neural networks. They assign probability distributions to the network weights, allowing for uncertainty estimation and more robust predictions. BNNs can be trained using techniques like variational inference or Markov Chain Monte Carlo (MCMC) sampling.
- **Sequential Monte Carlo methods:** these are also known as particle filters, Bayesian-based models used for state estimation and tracking in dynamic systems. They represent the probability distribution using a set of particles and update the distribution as new observations arrive.
- **Bayesian reinforcement learning:** these combine reinforcement learning techniques with Bayesian inference. It allows for incorporating prior knowledge about the environment and policies, enables uncertainty estimation, and provides a principled approach to exploration-exploitation trade offs.

Bayesian-based models offer advantages such as the ability to handle uncertainty, incorporate prior knowledge, update beliefs with new evidence, and provide probabilistic interpretations. They find applications in various domains, including natural language processing, computer vision, and decision making under uncertainty. A mutual fund selection example would be to select a fund based on sector preference by examining sector rotations and their impact on holdings in mutual funds. This could help to pick funds that are resilient to some macro shocks.

#### 4.2.4 MACHINE LEARNING-BASED MODELS

Machine learning-based models can learn complex patterns and relationships in the data that cannot be captured by linear regression models. These models can handle non-linear relationships between variables and adapt to complex decision boundaries, and can automatically learn relevant features from raw data, reducing the need for manual feature engineering. They are designed to handle large datasets with high computational efficiency and can scale well. They often employ ensemble methods, such as random forests or gradient boosting, to combine multiple models and improve overall performance. Their strength is in automatically selecting relevant features and identifying the most informative variables for the task at hand.

- **Decision trees:** these are hierarchical structures that recursively split the data based on input features to make predictions. They are easy to interpret and widely used in various applications.
- **Random forests:** this is an ensemble learning method that constructs multiple decision trees to make predictions.
- **XGBoost:** this is a “gradient” boosting algorithm that combines weak learners into a strong predictive model.
- **Support vector machines (SVM):** these are supervised learning models that classify data by finding optimal hyperplanes.
- **(e) Naive Bayes:** these are probability-based classifiers assuming independence between features given the class.
- **(f) Gaussian mixture:** these are probabilistic models that assume the data is generated by a mixture of Gaussian distributions. They are often used for clustering and density estimation.
- **(g) Hidden Markov:** these are statistical models that can capture temporal dependencies in sequential data. They are commonly used in speech recognition, natural language processing, and bioinformatics.
- **(h) Logistic regression:** this is a statistical machine-learning model used for binary classification problems. It estimates the probability of a binary outcome based on input features using a logistic function.
- **Principal component analysis (PCA):** this is a dimensionality reduction technique that identifies the most important features or patterns in data. It transforms the data into a lower-dimensional space while retaining as much information as possible.

Machine learning techniques encompass unsupervised and semi-supervised learning approaches, which can discover patterns and structures in the data without relying on explicit labels. These methods can be valuable for exploratory analysis, clustering, anomaly detection, and identifying hidden patterns. They are often designed to adapt and learn from new data, allowing them to handle changing environments. A mutual fund selection example would be utilizing a supervised machine learning algorithm like logistic regression to classify mutual funds. For example, finding the probability that a group of mutual funds with good historical performance would continue to have good future performance.

#### 4.2.5 DEEP LEARNING-BASED MODELS

These models have a human-like ability to learn based on non-linear and more complex relationships embedded in the data. Deep neural networks can automatically learn hierarchical representations of the data. They consist of multiple layers of interconnected nodes (neurons) that learn increasingly complex features at each layer. They can model non-linear relationships and capture complex patterns in the data. They can scale effectively to large datasets and are designed to handle big data scenarios, and can benefit from parallel computing on GPUs or distributed systems. This scalability allows for training models on vast amounts of data, which can improve performance and generalization. They often benefit from transfer learning, where models trained on large datasets or related tasks can be utilized as a starting point for new tasks. For example, pre-trained models, such as those trained on ImageNet for image recognition, offer a head start by leveraging prior knowledge and learned representations, reducing the need for extensive training on new datasets. They can extract relevant features from raw data automatically. Through multiple layers of abstraction, they learn representations that are useful for the given task. This feature extraction and representation learning make deep learning models effective in tasks such as image classification, speech recognition, and NLP. The disadvantages of “deep learning models” are that interpretability is challenging, and in general, they are data hungry, which means they require much more data for learning and to avoid overfitting.

##### 4.2.5.1 Early generation models

Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are particularly effective in capturing spatial and temporal patterns in these domains. CNNs are designed to capture spatial patterns and structures, while RNNs are effective in modeling sequential or time-series data. This makes them well-suited for tasks like object recognition, sentiment analysis, speech recognition, and machine translation. They can learn directly from raw input to output without relying on manual feature engineering or intermediate representations. This end-to-end learning simplifies the modeling pipeline and reduces the need for domain-specific knowledge and handcrafted features. They have demonstrated state-of-the-art performance on various tasks, surpassing traditional machine-learning approaches. They can achieve higher accuracy and better generalization, especially when trained on large-scale datasets.

- **Feedforward neural networks (FNNs):** these are inspired by the structure of the human brain; FNNs consist of interconnected nodes (neurons) organized into layers. They are usually used for reverse engineering or one-to-one mapping.
- **Convolutional neural networks (CNNs):** these employ convolutional layers to extract features from input data and are used for image recognition tasks.
- **Autoencoders (AEs):** these models are used for embedding and dimensionality reduction.
- **Recurrent neural networks (RNNs):** these are suitable for sequential data; RNNs utilize recurrent connections to capture temporal dependencies. Long short-term memory (LSTM) and GRU (gated recurrent unit) are also variations of recurrent neural networks that were introduced to help with vanishing gradients to avoid premature optimization.
- **Transformer models:** this is a type of architecture used for various tasks, especially natural language processing, due to their attention mechanism's ability to handle long-range dependencies effectively (particularly effective for natural language processing tasks, transformer models leverage attention mechanisms for sequence modeling).

#### 4.2.5.2 New generation models

Deep learning models benefit from ongoing research and advancements in the field. With the growing popularity of deep learning, new architectures, regularization techniques, optimization algorithms, and network designs continue to emerge, pushing the boundaries of what is possible. Recent advances in deep learning models are to work on human languages.

- **Generative models:** even though they may classify as part of deep learning models, we set them under their own due to their architecture and training. These models aim to generate new data instances that resemble the training data. Examples include: "generative adversarial networks" (GANs), which is a type of autoencoder that learns a probabilistic representation of data, enabling the generation of new samples; and "variational autoencoders" (VAEs), which is a type of autoencoder that learns a probabilistic representation of data, enabling the generation of new samples.
- **Transformer learning models:** these models leverage knowledge learned from one task to improve performance on a different but related task. Pretrained models like BERT for NLP or ImageNet-pretrained CNNs are common examples. Other examples include: "large language models" (LLMs), which are recent advances in deep learning models to work on human languages (the transformer architecture is the fundamental building block of all LLMs); and "generative pre-training transformer" (GPT), which is a language model that is pre-trained on sample data (tokens) to understand and then create language results, for example, for sentiment analysis.

In mutual fund selection, the methods and models often struggle to capture the complex patterns and stylized facts, potentially leading to suboptimal decisions. Generative adversarial models (GANs) or variational autoencoders (VAEs) can generate synthetic data that closely mimics the characteristics of real mutual fund data for better and more robust training of models.

### 4.2.6 HYBRID MODELS

Hybrid models refer to the combination of two or more different AI techniques or algorithms to create a single, more powerful, and effective model. For instance, reinforcement learning with deep neural networks (deep reinforcement learning) has been used in various applications. These models leverage the strengths of each individual technique, compensating for their weaknesses and improving overall performance. Hybrid models are often used to solve complex problems that may be challenging for a single AI approach to handle on its own. They can help make more informed decisions by combining different types of data, models, or strategies.

In mutual fund selection, they can (a) combine structured data (e.g., financial statements, price history) with unstructured data to gain a more comprehensive view of the fund's potential. By integrating various data sources, the model can identify patterns and relationships that individual models might miss and (b) help evaluate risks by combining traditional statistical models with machine learning algorithms. The statistical models may provide a solid foundation for risk estimation, while machine learning models can add the capability to analyze complex patterns and market dynamics, (c) use NLP techniques to analyze sentiment from news articles, social media, and financial reports, then combine the sentiment scores with other financial indicators to make more informed investment decisions, and (d) be designed to learn and adapt over time by combining reinforcement learning with other algorithms to continuously improve their decision-making abilities as market conditions change.

## 4.3 Training the models

Models need to base their decisions on some form of prior behavior that is set as an objective, and those suggestions can be accepted or rejected (including rewards or penalties for the more advanced models) at evaluation. Training techniques center around how to make the models learn the logic for making the suggestions. Feature engineering to extract meaningful patterns or relationships from raw data, which can help the model better understand the underlying patterns that can assure accurate classifications or predictions. This involves baseline training, testing (on out of sample data), and validation (during training to make sure there is no overfitting).

### 4.3.1 SUPERVISED TRAINING

The models learn from labeled training data to make predictions or classifications. They are provided with input-output pairs during training and aim to generalize patterns in the data to make accurate predictions on new, unseen data. Models learn to make predictions by minimizing the discrepancy between predicted and true labels. As such, labeled data means that for any input, the corresponding output is called a label, where input features are paired with corresponding target labels.

### 4.3.2 UNSUPERVISED LEARNING

The models learn by finding patterns and relationships in unlabeled data. They do not have explicit target labels during training with an aim to discover patterns, structures, or representations without explicit target labels. Unsupervised techniques often perform tasks like clustering, anomaly detection, dimensionality reduction, and generative modeling.

### 4.3.3 SEMI-SUPERVISED LEARNING

The models learn by a combination of supervised and unsupervised learning, where the model is trained on a small amount of labeled data and a larger amount of unlabeled data.

### 4.3.4 REINFORCEMENT LEARNING

The models learn by utilizing an agent interacting with an environment, learning optimal actions through trial and error. They receive feedback in the form of rewards or penalties for their actions and aim to maximize the cumulative reward over time. Reinforcement signals (rewards) guide the agent toward desired behavior.

## 4.4 Tuning the models

Training a model refers to the process of feeding labeled data into a model and adjusting its (internal) parameters so that it can learn to make accurate classifications or predictions on new (unseen) data. Tuning a model is the process of optimizing the hyperparameters of the trained model to improve its performance. These are parameters that are not learned during training but affect the learning process and the model parameters that result from it.

### 4.4.1 HYPERPARAMETER SEARCH

Hyperparameter search plays a vital role in fine-tuning machine learning models in order to do optimal performance. Grid search, random search, and Bayesian optimization are three common methods used for this purpose, each offering unique advantages. The choice of the hyperparameter search method depends on the complexity of the model, available computational resources, and the size of the hyperparameter space. By selecting the most suitable hyperparameters, machine learning models can deliver more accurate and reliable predictions for a wide range of real-world applications.

### 4.4.2 HYPERPARAMETER TUNING

Hyperparameter tuning involves optimizing model performance by fine-tuning hyperparameters, such as learning rate, regularization strength, batch size, and more. Unlike model parameters that are learned during training, hyperparameters are set before training and could significantly influence how the model learns and generalizes from data. Proper hyperparameter tuning is essential for achieving optimal model performance and preventing issues like overfitting or underfitting. By systematically adjusting these hyperparameters, learning models can better adapt to complex datasets and deliver more accurate and reliable predictions.

### 4.4.3 ONLINE LEARNING

Models are trained incrementally on streaming data, adapting to new information in real time. Particularly useful when data arrives sequentially or when computational resources are limited. Instead of waiting to accumulate a large batch of data and then retraining the model periodically, the online learning approach processes the data as it arrives. This approach enhances the ability to optimize outcomes that need real time assessments.

#### 4.4.4 AUTOML (AUTOMATED MACHINE LEARNING)

Automates the process of model selection (architecture), hyperparameter tuning, and feature engineering. Reduces the need for manual intervention, making AI more accessible to non-experts. For example, AutoML can automatically generate features (e.g., technical indicators, fundamental analysis metrics) that are relevant to predicting the performance of the mutual funds. By doing this, the system can sift through numerous holdings, analyzing different features, and can identify patterns and relationships that humans might overlook.

## 4.5 Assessing the models

Assessment techniques for AI models involve evaluating their performance, accuracy, and generalization capabilities. These assessment techniques provide insights into an AI model's performance, help identify areas for improvement, and ensure its suitability for the intended task or application. The choice of assessment techniques depends on the specific problem, type of model, and available data.

#### 4.5.1 ACCURACY AND LOSS METRICS

These metrics measure the model's performance on a specific task. For classification problems, metrics like accuracy, precision, recall, and F1-score are used. For regression problems, metrics such as mean squared error (MSE) or mean absolute error (MAE) are commonly used. These metrics provide quantitative measures of how well the model is performing.

#### 4.5.2 CROSS-VALIDATION

Cross-validation is a technique used to assess a model's performance by splitting the dataset into multiple subsets or folds. The model is trained and evaluated on different subsets, allowing for a more robust evaluation of its performance. Common cross-validation methods include k-fold cross-validation and stratified cross-validation.

#### 4.5.3 CONFUSION MATRIX

A confusion matrix is a table that summarizes the performance of a classification model by displaying the counts of true positive, true negative, false positive, and false negative predictions. It provides insights into the model's ability to correctly classify different classes and identify errors or misclassifications.

### 4.5.4 Receiver operating characteristic (ROC) curve and area under the curve (AUC)

These techniques are primarily used for binary classification problems. The ROC curve plots the true positive rate against the false positive rate at various classification thresholds. The AUC represents the area under the ROC curve and provides a measure of the model's ability to distinguish between classes.

#### 4.5.5 PRECISION-RECALL CURVE

The precision-recall curve illustrates the trade off between precision (the proportion of true positives among predicted positives) and recall (the proportion of true positives identified). It is particularly useful when dealing with imbalanced datasets or problems where one class is of greater interest than the other.

Overall, simply saying there is "AI" or transplanting solutions may not work as they have to be selected, tuned, trained, and refined for the tasks. To put this in perspective of our mutual fund selection example, assume we want to select five mutual funds (from a choice of thousands) given an objective (and evaluation criteria) of generating three-year excess return over the S&P500. Much like the multiple mutual funds, depending on the deployer's knowledge bank, there can be multiple models capable of dealing with multidimensional and temporal financial market datasets. These could include simplistic ones based on a single measure, a fixed equation, regression based, machine learning based, deep learning based, AutoML models, and so on. Furthermore, the circumstances themselves need to be modeled, including the interplay of measures, regimes, events, signals, sentiments, factors, etc. Another set of models could be for back-testing, where a model stability framework needs to be set up to continually assess if the chosen model is behaving the way it is supposed to and the triggers to note if/when the model is misbehaving and what to do (hopefully as a leading indicator). Simulation models can give color on the behavior of the selection under different scenarios. All this is in a continual loop of selection, evaluation of the variance from the objective, and refinement. It should also be noted that as we go beyond selection, other sets of models come into play, such as for asset allocation, portfolio management, risk management, asset planning, and so on.

On the change management side, this also highlights the need for robust "model risk management" (MRM) frameworks, especially for high-risk decisions, including:

- model change policies addressing periodic recalibration, data acquisition, algorithm decision overrides, dataset shifts, and replacement criteria
- using multiple shadow AI models, as recommended by regulators, to challenge and monitor the performance of the primary model
- establishing validation and audit standards.

## 5. VISUALS

Visuals: the appeal/taste of the dish remains a critical pillar. We believe visuals serve an important purpose in helping build trust around the analyses (however simple or complex). Without going on a psychology tangent, let us assume that human acceptance of results requires some degree of comfort around what the recommendation is for, when it is being made, and why it is being made. From a human-and-human perspective, this resides in the form of trust built around direct or implied relationships, needs, experiences, and so on. If we were to assume that human-and-AI interaction is also loosely based on a similar setup, then there needs to be a similar trust system. AI deployments attempt to build that trust by a) being accretive to some expectations (e.g., reduce time/effort, be profitable, etc.), and b) presenting them, at least initially, in a humanly digestible way (e.g., numerical and graphical representations that are appropriate, pertinent, experiential, etc.). This engagement is likely the key to accepted deployment, and as they say, a picture should speak a thousand words, or, in this case, become the face of the computational engines. We will not delve into the myriad of visual/presentation choices; simply put, if the visuals are not meaningful, intuitive, and easily explainable, then no matter how good the results, they may not be "useful" and will possibly be put in a drawer somewhere. In our opinion, for AI development to be trusted, it needs to be able to clearly represent the "what, why, and when" in a transparent and simple manner.

### 5.1 The what

The telos or the purpose of the AI application deployment. Holistically and locally, what is the purpose of the deployment? Is it accuracy, personalization, removing biases, eliminating emotions, supplementing information, expanding knowledge, automation, scaling analysis, remote or distant delivery, increasing solution points, increasing speed, reducing costs, increasing profitability, removing blind spots, identifying embedded relationships, recognizing patterns, detecting anomalies, faster execution, etc.? Yes, the choices and objectives can be very diverse and multiple, but they need to be articulated, understood, and set. The what, or the objective, is the key and is managed via accept-reject (or with reinforcement methods leveraging penalties and rewards) decisions in the training of the models. Herein, unless the telos or the overall objective is agreed to clearly, it may be a difficult deployment as the AI decision systems can be geared towards very different answers. AI deployment allows users to move from rule-based to decision-based ecosystems, but we note that these decision-based systems reside somewhat within rule-based ecosystems as critical decisions on objectives and judgments are arguably disguised rules with levels of granularity. And these need to be set.

In our mutual fund example, the objective can be to maximize excess return, and the evaluation can be to have a high Sharpe ratio. If the objective is to have a high return and low volatility, then you can set one as the primary objective (high return) and the other as a constraint (lower volatility), or use the "explainability index" approach to accommodate both. From a visuals perspective, setting the objective allows for easier visual representation for deployment, with numerical or graphical representations ranging from simplistic two-dimensional ones, such as line charts, to complex multidimensional ones, such as heatmaps, bar charts, stacked area charts, bar plots, parallel coordinate plots, etc. All with the motivation of providing data points to instill confidence and trust.

### 5.2 The why

This can be viewed as supporting representations for the decisions being "suggested". Any form of a model (whether complex or naive) can be a black box, depending on the user's sophistication. To build trust, we need to know why the decision is being made. However, given the accept-reject frameworks (or the penalty-reward functions for the more advanced models) embedded in the AI designs, it is easier to know if/when the outcomes are reliable than to know why (or how) the models are making them. And, as expected, increasing model complexity makes it exponentially more difficult to identify decision rationales. As a result, a lot of focus is on back-testing and simulations to test the models, but "the why" remains possibly the least structured part of the AI deployment processes. With increasing AI deployment, some methods are being suggested that try to explain the model's workings, including "partial dependence plots" (PDP), "permutation importance", "global surrogate models", and "anchors". Here we discuss the more common ones.

### 5.2.1 INTERPRETABILITY

This is to understand the relationship between elements in terms of the cause and effect (e.g., inputs and outputs); the drivers within the relationship for understanding the causality.

- **LIME (local interpretable model-agnostic explanations):** a technique for explaining individual predictions of black-box models. Generates locally interpretable explanations by perturbing input data and observing the impact on model predictions. Visualization tools can display the explanations, such as highlighting important regions in an image or showing word importance in text data.
- **SHAP (SHapley Additive exPLanation):** a method to attribute the contribution of each feature to the prediction outcome based on cooperative game theory.

### 5.2.2 EXPLAINABILITY

Understanding what is implied by the elements (e.g., inputs or outputs) in terms of what they all represent as part or whole. This facilitates data comprehension.

- **Attributions/contributions:** visualizing the attribution or contribution of input features to the model's predictions.
- **Feature visualization:** technique that is used to understand what features or patterns in the input data activate specific neurons in an artificial neural network.
- **Explainability index (EI) and "risk of target" (RoT):** technique that explicitly balances hundreds of input categories of performance measures according to default or specified preferences for a composite bounded score between 0 and 1 for each and the aggregate of the measures. RoT leverages the EI for comparing individual performance against benchmarks (as targets). The composite and component analyses explain the drivers of divergence of the target/objective (as a point-in-time, trend, or relative assessment). In the mutual fund example, this can also be used for managing multiple objectives and for reinforcements.

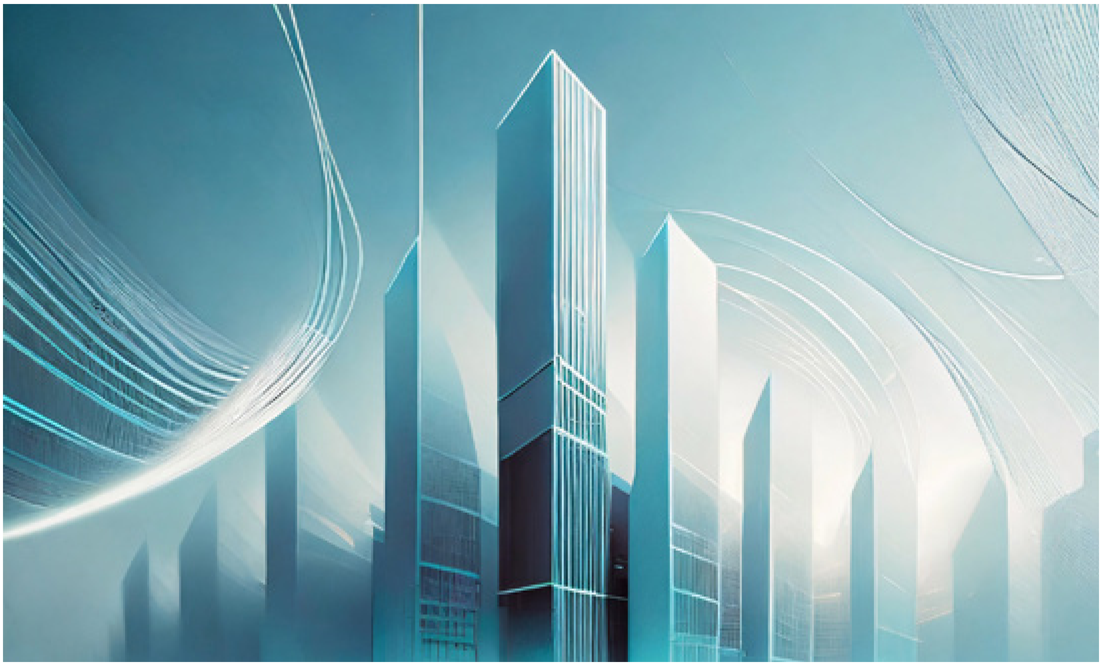


Image generated by Adobe Firefly

Regulatory concerns have added impetus to explainability and interpretability research. For example, under the E.U. GDPR, consumers are entitled to explanations for algorithm-driven decisions, a right not explicitly confirmed in the U.S. While AI may excel in credit scoring statistically, few E.U. banks seemed to have sought licenses for AI in internal credit evaluations due to regulatory concerns. In contrast, unregulated credit rating agencies heavily rely on AI.

### 5.3 The when

This can be viewed as the representations for the time period(s) being assessed. They can be absolute or relative and assess results as point-in-time or trends.

#### 5.3.1 Historical

Visualizing past data and trends to gain insights into historical patterns and relationships. Time series plots, line charts, and heat maps are commonly used for visualizing historical data.

#### 5.3.2 Prediction

Visualizing model predictions to understand patterns, trends, and potential future outcomes. Scatter plots, bar charts, and interactive visualizations are used to represent prediction results.

#### 5.3.3 Scenario analysis

Creating visual representations of hypothetical scenarios to explore the potential impact of different variables or events. Helps in decision making, risk assessment, and planning by visualizing various outcomes.

#### 5.3.4 Simulation

Visualizing simulations of complex systems or processes. Allowing users to observe and analyze the behavior of the simulated results. Graphs, animations, and 3D visualizations are common techniques used in simulation visualization.

#### 5.3.5 Back-testing

Visualizing the results during discrete points in time on out of sample or historic datasets. Helps in visualizing results during similar periods.

## 6. CONCLUSION

Our aim with the paper is to give the reader an appreciation of the multitude of ways to connect the dots, choices within use case deployments, possible variations in results, need for localized knowledge, dangers in oversimplification, need for cross sectional expertise, and so on. For the boxed cases, we may be more comfortable in pushing the proverbial deployment button (e.g., via a xxxGPT), but as the risks associated with the decisions increase, the deployment need and analysis may move across the quadrants, where understanding the nuances becomes critical in enabling optimum outcomes. As you read AI publications, you will note that the AI deployment itself is no different. The preference is somewhat in the eye of the beholder and pitched around the deployer's knowledge (including searchable methods) that is influenced by their backgrounds and agenda, e.g., economist, mathematician, philosopher, politician, etc. For example, economists tend to lean on the cost or value wrappers, and philosophers on the choice wrappers.

One way to think of the deployment optionality spectrum is as a range from acceptable imperfection (i.e., with lower accuracy, higher error tolerance, low efficiency per training data, weak models, weak infrastructure, etc.) to assumed perfection (i.e., with generally reduced choices with hyper personalization). We note that a) perfection itself is transitory, as most methods are based on available knowledge banks that are rapidly evolving, and b) current AI deployments can largely only handle accept-reject functions (with degrees of reward-penalty functions for more advanced ones); they are weak in managing the grayer human aspects such as implied meaning, emotions, evolving expectations, intentions, gut, valuing collateral damage, etc. The question becomes how is the telos (or, for that matter, your thinking) placed on the spectrum? Going to our food analogy, just because we know certain food types are not good, do you entirely stop eating them? Do you only go to the very "best" restaurants? How much of the freedom of choice can you give up? How quickly do you cede control to the "suggestions"? These decisions are easier for some tasks than others. As you frame answers to these choices, you start forming your deployment spectrum placement and path.

Since AI deployment is technical, the question regarding whether you need supplemental expertise and from whom arises. Experts are putting stakes in the ground with publications and packaged solutions. Incumbents with legacy infrastructure and capital investments have varying degrees of inertia and appetite for discovery. New entrants' nimbleness allows for speedy delivery but generally comes with a higher focus on beauty and experience, so the cut/paste of models becomes risky as the results can be very questionable. Either way, not everyone can engage in the advanced quadrants, as that requires knowledge, time, and capital. Herein is our word of caution, the race to AI everywhere that is now being accentuated with the xxxGPT claimants across verticals has dangerous elements, especially when combined with the traditional tech industry mindset of accepted risk of failure in getting the minimal viable product out. Maybe herd decisions will make some use cases subject to self-fulfilling prophecies, but where the risks associated with deployment are high you need to be cautious.

In navigating these elements lies the key to mitigating adverse surprises akin to the Y2K and the then some money burning adventures. We believe effective AI deployment lies in the knowledge intersection of subject matter, computer science, data science, and machine learning expertise. Advanced users understand the importance of what is under the hood and casual users base the usage on trust, which is earned. Either way, we find that meaningful AI deployments demand more than a simplistic “data in, miracles out” strategy. They require meticulous tuning, enhancements, and occasionally rethinking approaches. As such, great experiences blend the exterior with calibrated power under the hood.

# APPLIED GENERATIVE AI GOVERNANCE: A VIABLE MODEL THROUGH CONTROL AUTOMATION


**GERHARDT SCRIVEN** | Managing Principal, Capco[1]

**MARCEL BRAGA** | Principal Consultant, Capco

**DIOGO SANTOS** | Principal Consultant, Capco

**DIEGO SARAI** | Managing Principal, Capco



## ABSTRACT

Generative AI has the potential to revolutionize the banking industry with hyper-personalization and advanced chatbots. However, the technology also poses risks to trust, accuracy, fairness, intellectual property, and confidentiality that all need to be mitigated to ensure that the benefits of Generative AI are realized. In this article, we explore practical considerations to help mitigate these risks through the construction of a governance framework that has a focus on AI explainability, intellectual property protection, and minimizing model hallucination. We then derive a control framework against these key outcomes and present technology solutions we built around automating some of the key controls towards making our governance model viable. Finally, we explore what other institutions are doing in the field of generative AI governance and discuss new emerging roles needed to execute against the governance model. In terms of practical application, we recommend that financial institutions start small when it comes to generative AI governance and focus on defining a "minimum governance model" on a use case by use case basis to minimize the time and cost footprint of governance. We also recommend that governance is implemented very early in the solution lifecycle so that it is baked in at root-level; hence, reducing churn and rework of the solution when industrializing the use case within the financial institution.


## 1. INTRODUCTION

Generative AI has the potential to revolutionize the banking industry, from a business as well as a technology perspective, by enabling hyper-personalization around financial planning, investment portfolios, product recommendations, and financial education. Moreover, personalized customer service can be provided to clients using the technology through advanced chatbots that provide tailored responses based on the customer's financial history and preferences.

However, as AI systems become more advanced and integrated into the banking industry, there is a growing need to understand and manage AI-related risks to ensure that the benefits of AI are realized while potential negative consequences are minimized.

More precisely, generative AI, which encompasses techniques such as deep learning and generative adversarial networks and include "large language models" (LLMs – generative AI that specializes in text understanding and generation), has the potential to create highly realistic and sophisticated outputs, including fake information and malicious code. This poses a range of risks, such as erosion of trust in financial institutions and the risk that AI may provide sub-par or incorrect recommendations and advice to bank personnel or the financial institution's customers.

Additional generative AI risks that financial institutions need to be aware of, and mitigate, include:


[1] We would like to thank the sponsors of this work: Alessandro Corsi and Luciano Sobral.

- **Bias and fairness:** generative AI models can inherit and perpetuate biases present in their training data, leading to biased content generation and reinforcing existing inequalities.
- **Intellectual property infringement:** generative AI models can generate content that infringes upon copyrights and trademarks, posing legal challenges.
- **Data protection:** to obtain the best results from generative AI for specialized tasks, it is often necessary to finetune the AI models with contextual information pertaining to the knowledge domain the solution will address, either in the form of training or via prompt engineering (crafting input instructions or queries to achieve desired outcomes when using Gen-AI). Herein lies an additional potential risk, that of protecting corporate intellectual capital as well as personal information of customers. For the latter, generative AI systems that process personal data must be designed in a way that protects the privacy of that data. This includes implementing appropriate security measures and providing individuals with control over their personal data, as stipulated by regulations such as the European Union's General Data Protection Regulation (GDPR).

As a secondary driver for governance: as one builds out automated solutions around addressing some of these risks, one also needs to be sure that one can trust the automated processes.

To create a holistic approach for managing AI risk, Tan (2023) presents a "generic AI risk management framework",[2] which consists of six pillars, of which "governance & oversight" is a key component to manage the other five pillars.

In this paper, we will explore the governance pillar in more depth and focus on the practical considerations (applied AI governance, a corresponding control framework, and emerging roles needed to manage generative AI) in order to help mitigate AI-related risks. The topics covered here will be particularly relevant for readers who are relatively new to implementing solutions using generative AI technology in corporate environments.

**Figure 1:** A generic AI risk management framework


Source: Tan (2023)

## 2. DEFINING A GENERATIVE AI GOVERNANCE MODEL

In establishing any governance model, a good starting point is to define the desired outcomes that one wants to achieve through applying the model. In the case of generative AI, there are three key outcomes that need to be considered.[3]

### 2.1 Being able to explain the results from AI

It is critical to be able to explain how AI, and in particular generative AI, arrived at a certain result.

- **Transparency:** explainability allows stakeholders to understand and trace how the AI system arrived at its conclusions or generated its outputs. This helps build trust and confidence in the technology.
- **Bias detection and mitigation:** explainability enables the identification of biases or unfairness in the AI system's outputs. By understanding the underlying processes and decision making, biases can be detected and addressed, leading to fairer and more equitable outcomes.

[2] https://tinyurl.com/27898j48
[3] It should be noted that the three outcomes discussed below are not exhaustive and that there are other dimensions of generative AI governance. Others include protecting AI models from adversarial attacks and ensuring that AI models are performant and scalable. However, these are well established AI-related governance topics, whereas the three key items listed below require new or significant additional thinking specifically for generative AI.

- **Error detection and correction:** explanation capabilities help identify errors or mistakes made by the AI system. Users can understand why certain outputs may be incorrect or undesirable, allowing for improvements and corrections to be made.
- **Intellectual property and ownership:** Explainability can help establish ownership and intellectual property rights in AI-generated works. By understanding the creative process behind AI-generated content, individuals and organizations can assert ownership and defend their rights.

## 2.2 Protecting intellectual property and sensitive information

AI governance can help to protect corporate intellectual property by ensuring that it is properly identified and managed during the information processing lifecycle.

- **Protecting sensitive information:** AI models are often trained on sensitive data, such as customer data and financial data. By protecting the confidentiality of this data, organizations can avoid high-impact risks, such as data breaches, reputational damage, and regulatory fines.
- **To comply with regulations:** many regulations require organizations to protect intellectual property (IP) and sensitive information about their customers and employees. For example, the General Data Protection Regulation (GDPR) requires organizations to implement appropriate security measures to protect personal data.

## 2.3 Combating hallucination

Generative AI models can be used to create content that is sometimes indistinguishable from real content (hallucinate), which can lead to people being misled or deceived. When this happens, trust in AI and the institution that served the content can be undermined. Combating hallucination is hence

**Table 1:** Controls for the three key generative AI governance outcomes

| KEY OUTCOMES OF OUR GENERATIVE AI GOVERNANCE MODEL | | |
|---|---|---|
| BEING ABLE TO EXPLAIN THE RESULTS THAT AI PROVIDED | PROTECTING INTELLECTUAL PROPERTY AND SENSITIVE INFORMATION | COMBATING HALLUCINATION |
| **Corresponding controls that are used to determine whether outcomes are being achieved** | | |
| 1. There is a clear traceable connection between input (provided to AI as context) and the result returned by AI.<br>2. The AI system employs explainable AI techniques to provide interpretable explanations for its decisions.<br>3. The AI model is validated using specific transparency metrics to ensure its decision-making process is transparent.<br>4. The explanations provided by the AI system are audited by third-party experts to verify their accuracy.<br>5. Multiple AI models are employed to provide insights into the decision-making process, factors they consider, and the explanations they provide for their outputs.<br>6. Regular fairness/bias testing cycles are conducted.<br>7. Fairness-aware algorithms are employed during model training. | 1. Any data sent to generative AI models is thoroughly vetted to ensure that it does not contain sensitive or proprietary information. Specifically, data minimization techniques are employed (only provide the LLM with the minimum amount of sensitive corporate material necessary for its intended purpose).<br>2. Data anonymization and redaction techniques are employed to remove any identifying information from the input data.<br>3. The generative AI model and information processing pipeline are deployed in an environment with restricted access, preventing unauthorized access to sensitive information.<br>4. Data usage audits are conducted regularly to verify compliance with intellectual property protection policies.<br>5. A data inventory is maintained.<br>6. Regular privacy audits to test for compliance are conducted. | 1. Guardrails are applied to the generative AI models to prevent them from providing information outside set boundaries.<br>2. Generative AI models are tested on a diverse set of inputs, including edge cases and outliers, to verify that they do not generate unrealistic or nonsensical outputs.<br>3. An ensemble of generative AI models is used to cross-validate outputs and reduce the risk of hallucination.<br>4. Adversarial testing is performed to assess the model's resilience against potential hallucinatory inputs.<br>5. Generative AI models are continuously monitored in production to detect any potential cases of hallucination.<br>6. A real-time alerting mechanism is in place to notify responsible personnel if the AI model breaches guardrails.<br>7. In the event of guardrail breaches, a well-defined incident response plan is activated to investigate, rectify, and prevent similar incidents in the future. |

**Table 2:** Application of the controls over the knowledge lifecycle

| | PRE-PROCESSING | IN-PROCESSING | POST-PROCESSING |
|---|---|---|---|
| | Pertains to the first analysis of raw information that will ultimately be used to finetune generative AI models. The approach involves annotating, or marking up, raw data that will facilitate tracing output from Gen-AI. | Pertains to filtering and redacting the augmented raw input information towards ensuring that generative AI models receive the smallest amount of information required to perform its tasks. | Pertains to tasks that need to be executed as part of testing and monitoring AI-models. This also includes proactive steps that can be taken to ensure generative AI behaves within set parameters and boundaries. |
| Controls related to being able to explain the results that AI provided | | | |
| Controls related to protecting intellectual property and sensitive information | | | |
| Controls related to combating hallucination | | | |

a critical governance objective and companies need to ensure the accuracy and reliability of Gen-AI models and work to build trust in AI. If end users cannot trust that generative AI models are producing accurate and reliable results, they are less likely to use them. This could hinder the adoption of the technology and realizing its potential benefits.

## 3. THE CONTROL FRAMEWORK

Defining a set of controls lies at the heart of any pragmatic governance model and represents the first step in building out our model. The controls define the mechanism by which one can comprehensively measure whether any given outcome is being achieved. A set of controls that can help achieve the three key outcomes described above is presented in Table 1.

It should be noted that corporate proprietary information and other sensitive data that are used to provide generative AI models with context towards assisting with specialized use cases generally go through a "knowledge lifecycle" that comprises of three steps: knowledge preparation (pre-processing), informing generative AI (in-processing), and knowledge consumption (post-processing). The controls that we defined in Table 1 should logically be applied during specific points across this lifecycle. Table 2 demonstrates this.

Applying all the controls listed in Table 1 is not a trivial task. To fully implement the generative AI governance model, processes need to be built around these controls so that they can execute and be reported against. To practically apply these controls, we strongly recommend maximizing automation around the supporting processes.

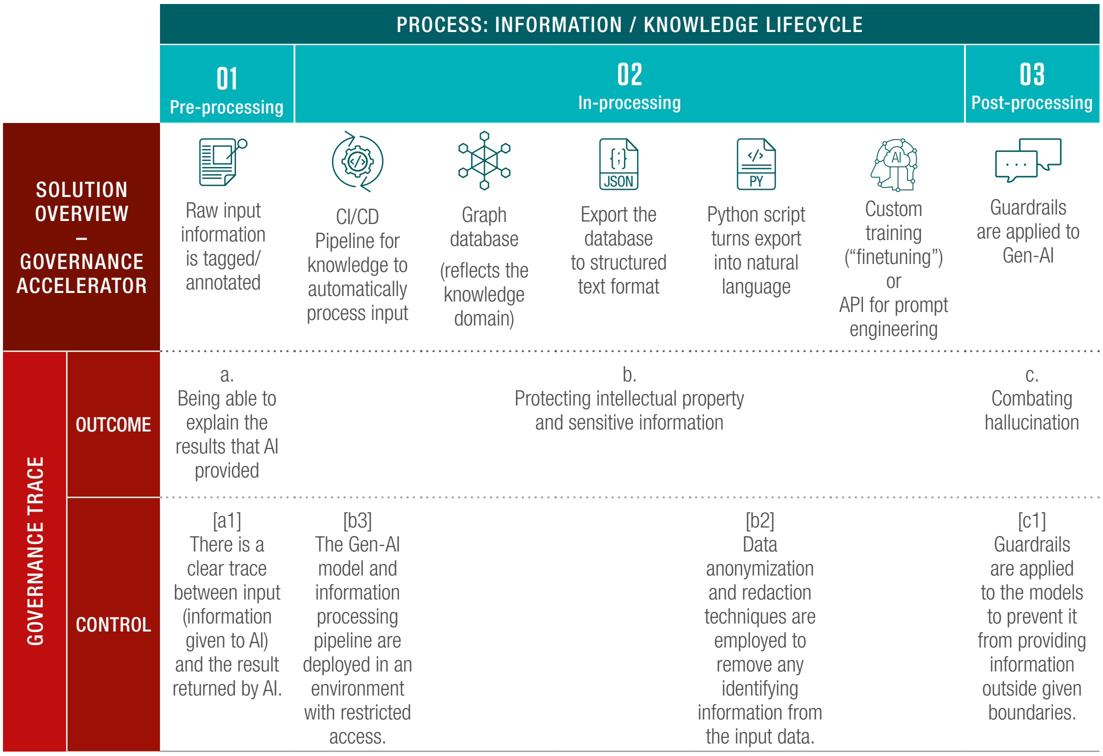
It should be mentioned, however, that while automation does not present a silver bullet towards AI-governance, it can significantly impact the cost and time footprint of executing many of the controls.

As part of our internal research, we have made significant advances in automating the execution of some of the control points listed in Table 1. These will be explored in the next section. We will also provide an overview of some of the work that other institutions are doing against some of the other control points we mentioned.

## 4. APPLIED AI-GOVERNANCE

Figure 2 provides an overview of a solution we have designed and built around the three "knowledge lifecycle" components we described earlier. The solution, which comprises of a collection of "control automation accelerators", traces to both the generative AI governance outcomes we defined as well as some of the control points that were listed in Table 1:

**Figure 2:** Generative AI governance automation solution overview




Our solution can be divided into three sub-solutions that directly correlate with the three key outcomes we described earlier:

### 4.1 Being able to explain the results that AI provides

Key to explaining AI's responses is to create a trace between generative AI's output and the material that was provided to it as input through finetuning.

To achieve this goal, we apply automated content markup and classification, through a couple of steps:

- We break the input knowledge that will be fed to AI into smaller fragments, such as pages or paragraphs (The raw input data may be in the form of large text-based documents, such as large PDF files, and be federated across multiple repositories).
- Metadata, in the form of keywords, are extracted from the fragments using RAKE (rapid automatic keyword extraction)[4] and synonyms of these keywords are obtained via consulting generative AI. This metadata, together with contextual information about the location (page number, paragraph number, information repository link, etc.) of the knowledge fragment is added to the input that is provided to AI.

During information retrieval:

- We dynamically match the user's query with the metadata we extracted and, towards better system performance, we only share content where we have a good match with generative AI.
- As part of this, we include the contextual information regarding where in the document the fragment comes from (page, paragraph, etc.).
- When AI responds to the user's query, it references the source by using a template to format generative AI's response in a way that includes the page and paragraph number.

[4] https://tinyurl.com/3xezt826

As a result, we can provide a detailed trace between the response and the specific information AI used to generate the response, which significantly facilitates explainability.

### 4.2 Protecting intellectual property and sensitive information

This outcome can be readily attained by using end-to-end automation in processing the information from its genesis point to where it is handed to AI as part of custom training or as prompt input. The key principle we apply here is that no human hands should touch the data.

To achieve this, we defined a solution that operates across the in-processing sub-process of the "information lifecycle process" (Figure 2):

- Harvests knowledge/information the moment it gets published into a version control system (such as Git or Subversion) through using a CI/CD (continuous integration/ continuous delivery) pipeline for assets solution. The knowledge base that is harvested can be of any type – images, videos, audio, 3D models, data files, etc.
- Automation ensures that this information is processed, and a knowledge graph (which models/reflects the underlying knowledge domain) is updated accordingly.
- The knowledge graph is periodically exported, and the export file is converted to natural language through using a Python script, and from there it is injected automatically into a large language model (LLM) by using the LLM's "application programming interface" (API).
- Checks and balances along the processing pipeline ensure that what is sent to the generative AI model is appropriately filtered, redacted, or anonymized.

Because the entire process is automated, and access to the data in any stage of processing is highly restricted, corporate intellectual capital and other sensitive information is much better protected.

#### 4.3 COMBATING HALLUCINATION

Hallucinations can be caused by a number of factors, such as the quality of the training data, the complexity of the model, and the way in which the model is used. To combat hallucination, we apply concepts we introduced earlier:

- **Use of external knowledge:** by incorporating contextual and external knowledge into the model, the likelihood of hallucination is reduced through providing AI with a more accurate representation of the world, within the context of the specific use case.
- **Data augmentation:** this technique involves transforming training data in various ways to expose the model to a wider range of patterns. By doing so, the model becomes more robust and less prone to hallucinating. In our practical example, we accomplished this by adding synonyms of key concepts that are addressed in source knowledge to the metadata that is used for prompting AI.

Additionally, we constrain AI through smart prompting (we script additional instructions and add it to the end user's input) to only employ the source knowledge we provided for constructing its responses. To do this, we take control of the entire user experience lifecycle and supplement user queries with these additional instructions in the background, i.e., explicit instructions to prevent AI from generating unrealistic or nonsensical responses. Finally, we set confidence thresholds for generated outputs. If the model's confidence falls below a certain threshold, the output can be flagged for further review or discarded to avoid potential hallucination.

## 5. A BRIEF OVERVIEW OF WHAT OTHERS ARE DOING

Several other institutions are conducting research in AI governance. We try and connect some of these endeavors to a subset of the controls we defined in Table 2.

### 5.1 Being able to explain the results that AI provides

#### 5.1.1 CONTROL #2

The AI system employs explainable AI techniques to provide interpretable explanations for its decisions.

- Google AI has developed a number of explainable AI techniques, including LIME and SHAP (Google Colab).[5] These techniques are used in a variety of Google products, such as Google Search and Google Photos.
- Microsoft Research has also developed a number of explainable AI techniques (Explainability – Microsoft Research). These techniques are used in Microsoft products, such as Microsoft Azure Machine Learning and Microsoft Power BI.

[5] https://tinyurl.com/3ywd2cdw

- IBM Research develops and applies explainable AI techniques to a variety of problems (Explainable AI | IBM Research),[7] such as fraud detection and healthcare decision making.
- Amazon Web Services offers a number of explainable AI services, such as Amazon SageMaker Explainable AI.[8]

### 5.1.2 CONTROL #3

The AI model is validated using specific transparency metrics to ensure its decision-making process is transparent.

To ensure that generative AI models are used transparently, it is important to define and then validate them using transparency metrics. Some of the key terms involved in LLM transparency include:

- **Perplexity:** a measure of how well an LLM can predict the next word in a sequence.
- **Coherence:** a measure of how well the LLM's outputs make sense semantically.
- **Context appropriateness:** a measure of how well the LLM's outputs are relevant to the given context.

According to AIMultiple,[9] one of the key steps that organizations can take to validate LLMs for transparency is to use multiple evaluation metrics. Instead of relying solely on perplexity, for example, incorporate various evaluation metrics that capture different aspects of the LLM's performance, such as the ones we listed above. Moreover, it is important to implement transparency by design. One approach is using the "community transformer" design, which is a type of LLM architecture that is designed to offer a higher level of transparency than traditional LLM architectures. This design specifically allows users to see how the LLM is attending to different parts of the input sequence and how it is making its predictions.

### 5.1.3 CONTROL #7

Fairness-aware algorithms are employed during model training. Cornell University reported the following regarding employing fairness-aware algorithms[10]:

Familiarize yourself with different fairness definitions and metrics to identify the most suitable ones for your specific Gen-AI application.

Some "fairness definitions" include (but are not limited to) "demographic parity", the proportion of individuals from different protected groups (e.g., gender, race, ethnicity) who receive a favorable outcome should be equal; "individual fairness", which states that individuals who are similar in all relevant respects should receive similar outcomes, regardless of their protected group membership; and "counterfactual fairness", which states that individuals should receive the same outcome that they would have received if their protected group membership had been different.

Some "fairness metrics" include (but are not limited to) "discrimination ratio", which is calculated by dividing the proportion of individuals from a protected group who receive a favorable outcome by the proportion of individuals from a non-protected group who receive a favorable outcome, and "fairness-aware accuracy", which is calculated by taking the weighted average of the accuracy for each protected group, where the weights are determined by the size of each protected group.

Fairness metrics should be applied across different moments in the information lifecycle, including:

- **Pre-processing:** pre-processing techniques must be applied before data is fed to generative AI. This can include re-sampling, re-weighting, or transforming the data to ensure a more balanced representation of different groups.
- **In-processing:** incorporate fairness-aware optimization techniques during the LLM training process. These techniques can help balance the trade-off between model accuracy and fairness by adjusting the model's parameters or loss function.
- **Post-processing:** refers to post-processing techniques used to adjust the model's outputs to ensure fairness. This can include thresholding or calibration methods to achieve desired fairness metrics.

It is important to compare different fairness-aware algorithms and techniques to identify the most effective approach for your specific use case.

[7] https://tinyurl.com/3tm7kakp
[8] https://tinyurl.com/ywf64vt3
[9] https://tinyurl.com/mwphnjfn
[10] https://tinyurl.com/bdd5h5wn

## 5.2 Protecting intellectual property and sensitive information

### 5.2.1 CONTROL #5

A data inventory is maintained, which identifies, collects, and organizes personal data in systems, tracks data sources, and helps map how an organization's data assets are stored and shared. Although the concept of a data inventory is not new, it has gained prominence in recent years due to regulations like GDPR and CCPA (California Consumer Privacy Act), which require companies to have greater control over their data and to help organizations identify sensitive data.

According to RedClover Advisors,[11] a "data inventory" solution is predicated around data collection, usage, storage, and sharing practices; types of data collected; who data has been collected from; whether the data falls into any sensitive categories; and consent requirements.

Within the context of Gen-AI, some of the challenges in creating an efficient data inventory for "large language models" (LLMs) include:

- **Complexity of datasets:** LLMs require large volumes of data for training, which can make organizing and managing this data challenging.
- **Timeliness of information:** LLMs may not have updated information, as their knowledge is based on the training data available at the time of training.
- **Data source integration:** injecting knowledge into LLMs from various sources, such as external structured databases or company-specific APIs, can be challenging.
- **Data fragmentation and silos:** the existence of data silos and fragmentation of information across different platforms and systems can hinder the creation of a comprehensive and efficient data inventory.

To overcome these challenges and facilitate the creation of a data inventory, the following actions can be followed: implement a data warehouse or data lake to store all data used for training AI in a centralized location, which will make it easier to create and maintain a comprehensive and efficient data inventory; use a data management platform, such as Apache Hive, to help organize and manage large volumes of data; use a data pipeline to automate the data lifecycle, such as the CI/CD pipeline for assets solution we described earlier, together with version control to track changes to the data inventory; and use a data integration platform, such as Apache Nifi, to connect data silos and fragmentated information across different platforms and systems.

## 5.3 Combating hallucination

### 5.3.1 CONTROL #3

An ensemble of generative AI models is used to cross-validate outputs and reduce the risk of hallucination.

Robust Intelligence presented an approach for using an ensemble of generative AI models to reduce the risk of hallucination through:[12]

- Choosing a variety of generative AI models with different architectures, training data, or hyperparameters, diversity in their predictions can be ensured to reduce the likelihood of all the models hallucinating in the same way.
- Combining model outputs by using techniques such as voting (for classification tasks) or averaging (for regression tasks).
- Evaluating ensemble performance by using metrics relevant to the specific application/use case. Comparing the ensemble's performance to that of individual models is also important for ensuring that the ensemble is providing improved results.

### 5.3.2 CONTROL #6

A real-time alerting mechanism is in place to notify responsible personnel if the AI model breaches guardrails. According to Tata Consulting Services, a real-time alerting mechanism that notifies responsible personnel if the LLM breaches guardrails can be implemented using the following steps:[13]

- Establish clear guardrails for generative AI, i.e., a set of programmable constraints and rules that monitor and dictate user interactions with the model, ensuring it operates within defined boundaries and adheres to specific rules or principles. Examples of such guardrails include: "accuracy guardrail", which ensures that the AI model is performing as expected and is meeting its accuracy goals (should the model's accuracy fall below the set value an alert will be sent out); and "bias guardrail", which ensures

[11] https://tinyurl.com/mr32xvb2
[12] https://tinyurl.com/r2m2uyrt
[13] https://tinyurl.com/4b4e5hmk

that the AI model is not biased against any particular group or individual (the fairness metrics we referenced earlier can be used to define thresholds, which, if violated, will trigger the alerting mechanism).

- Continuously monitor the LLM's performance and outputs in real time, checking for any breaches of the established guardrails.
- Develop an alerting system that triggers notifications to responsible personnel when a breach of guardrails is detected. Email, SMS, and Slack messages are examples of potential alert carriers.

## 6. EMERGING ROLES AND EXECUTIVE PARTICIPATION

To support new processes that need to be built around the controls that we defined in Table 1, new roles will need to emerge. Here are a few examples:

- **AI Governance Lead:** this role will oversee the implementation and execution of the generative AI governance model and control set. The Lead will need to have a deep understanding of AI technology and the risks and challenges associated with its use in the financial services industry.
- **AI Risk Manager:** this role will be responsible for identifying and assessing the risks associated with the use of generative AI, and thus have a strong understanding of general risk management principles and practices.
- **AI Compliance Officer:** this role will be responsible for ensuring that the use of generative AI complies with all applicable laws and regulations, and needs to have a strong understanding of the legal and regulatory landscape for AI in the financial services industry.
- **AI Ethics Officer:** this role will be responsible for ensuring that the use of generative AI is ethical and responsible.
- **AI Technical Architect:** this role will be responsible for designing and implementing the technical infrastructure to support the generative AI governance model, and needs to have a deep understanding of AI technology stack and the associated infrastructure requirements.

Moreover, the implications of generative AI governance for CIOs, CTOs, CFOs, and business leaders are also significant. CIOs will need to ensure that the IT infrastructure is in place to support the generative AI governance model. This includes providing the necessary computing resources, data storage, and security controls. CIOs will also need to work with other stakeholders to develop and implement policies and procedures for the responsible use of AI.

CTOs will need to work with the AI Governance Lead to ensure that the generative AI governance model is aligned with the overall IT strategy. CTOs are ultimately responsible for implementing and maintaining the generative AI governance model as well as developing and deploying the necessary tools and technologies.

CFOs will need to budget for the costs of implementing and maintaining the generative AI governance model. This includes the costs of new roles, as well as the costs of new tools and technologies.

Business leaders need to ensure that the AI governance model is effective in meeting the needs of the organization. This includes understanding the importance of generative AI governance and being comfortable that AI solutions are being used in a way that aligns with the financial institution's values and principles towards building trust with customers, employees, and regulators; being involved in the development and implementation of the generative AI governance model to ensure that it is aligned with the organization's overall business strategy; and helping with monitoring and evaluating the generative AI governance model on an ongoing basis, since it needs to be adapted to changes in the regulatory landscape.

## 7. CONCLUSION

Generative AI presents additional challenges in the domain of AI governance, particularly around key outcomes such as transparency, protection of sensitive information, and combating hallucination. Defining a lean set of controls that trace to the outcomes and building supporting processes around these controls are at the heart of establishing a pragmatic governance model.

In our research, we were successful in partially achieving the desired outcomes by applying a combination of control automation in the information processing lifecycle, together with techniques to better contain generative AI within a clear set of boundaries to combat hallucination. We previously reported on this in an earlier article.[14]

Moreover, the need for AI-related governance is well recognized in the industry and many institutions have provided solutions around some of the controls we discussed. The solutions referenced in this paper together with our own governance accelerators collectively form an excellent primer for establishing a robust generative AI governance practice within an enterprise.

In closing, some final points key points about generative AI governance are that:

- Much of it is new and complex.
- It can radically change ways of working and how reliability is assessed.
- It involves many and very disparate stakeholders.
- It goes to the heart of key processes (such as client interactions, delivery at quality).
- It is not a one-time event. It is an ongoing process that needs to be adapted continuously to changes in AI technology and the regulatory landscape.

Hence, it is important to pay attention to testing the governance model as one develops it, much in the same way that for a project one needs to test the governance and delivery methodology.

Towards this end, we have the following recommendations when establishing a generative AI governance model within a financial institution:

- Not all Gen-AI use cases will require the same level of governance and control. Hence, we recommend defining a "minimum viable governance" (MVG) model on a case-by-case basis.
- Define and implement the MVG when the use case is in pilot phase already. This is because retroactively applying a robust governance structure is likely to result in significant churn in the core solution.

[14] https://tinyurl.com/27kp9d5e

# AI AND BANKS. IN CONVERSATION WITH AN AINTERN

**JESÚS LOZANO BELIO** | Senior Manager, Digital Regulation, Regulation and Internal Control, BBVA

## ABSTRACT

The rapid advancement of artificial intelligence has facilitated the automation of previously challenging tasks. This article explores the opportunities and benefits associated with AI adoption, specifically within the banking sector. It examines how banks are currently utilizing AI, the challenges they face in implementing AI systems, and the role of regulators in supporting AI adoption. Additionally, as this article has been written with the help of some AI tools, it serves as a practical demonstration of AI's applicability in research and information dissemination. While AI demonstrates proficiency in these areas, it is important to note that human expertise and supervision remain essential due to inherent limitations of the technology.

## 1. INTRODUCTION

Ever since I joined the Digital Regulation team of BBVA in December 2016, where I was tasked with monitoring AI regulatory developments, I felt compelled to make use of AI for my work. Indeed, I coined the term "AIntern" to refer to a tool that was able to perform some of the tasks that an intern in a research department is usually entrusted with, such as compiling data and information, summarizing papers, and helping to write reports.

Unfortunately, I was not able to engage in such a project due to time and resource constraints and I gradually abandoned the idea. However, when generative AI hit the headlines at the beginning of 2023, that idea came back to me and I started to explore the different tools that are available in the market. My objective was to take advantage of these tools to build my own AIntern as soon as I had some spare time.

Sadly, 2023 has resulted in a very prolific year in digital regulation and I had to postpone my plans once again. However, when I was invited to write an article for the Journal of Financial Transformation, I immediately thought of how I could turn my contribution into a proof of concept of my long-awaited AIntern project.

Thus, this article is the outcome of that proof of concept. Its content has been written with the invaluable help of some AI tools that I will refer to in the appendix. The content itself is not revolutionary as it focuses on the opportunity of using AI, the benefits it can bring, how banks are using AI, the challenges linked to AI adoption, how regulators can support the adoption of AI by banks, and some basic steps to take into consideration for the successful adoption of AI.

This article is based on a combination of reports, news, and papers I have written or contributed to and other documents I have read since 2017 that I found interesting, and all of which are publicly available. I gave them to my AIntern and then questioned it about them. What follows is a summary of "the interview" I did with my AIntern.

In the appendix, some additional information on the questions I asked it, the sources used, and the process to select them, among others, can be found. It is my intention that this appendix helps readers identify some of the considerations that have to be taken into account when using generative AI[1] for these types of tasks.

[1] A brief explanation of this type of AI implementation can be found in McKinsey & Company (2023).

My main takeaway from this proof of concept is that these tools are mature enough to support humans in writing correct and generally accurate text in areas that the human has some previous knowledge of, and, therefore, is able to perform some editorial control. Nevertheless, as you will see in the next section, the responses are reasonably accurate after my "editorial intervention", but they lack the human touch an experienced human could deliver.

Nonetheless, the latter could also be an indication of my limitations as a user of AI applications or of the applications I have chosen and tried, or the need to devote more time in obtaining catchier answers, which would be to detriment of the time saved for producing these texts.

**Figure 1:** AIntern interview

Image created by the author with AI Picasso app (aipicasso.app) and Imgcreator (imgcreator.zmo.ai)

## 2. THE INTERVIEW

Below, I present some of the questions posed and the answers provided by my AIntern.

### 2.1 Why is this the right moment for AI adoption?

**The response from my AIntern was:** The current era is considered the perfect moment for the widespread adoption and application of AI due to three key factors:

- **Technological advancements:** AI technologies, particularly in machine learning and deep learning, have made significant strides in recent years. These advancements have greatly enhanced the capabilities of AI systems, enabling them to process and analyze large volumes of data, recognize complex patterns, and make accurate predictions.
- **Availability of data:** the exponential growth of digital data and the increasing connectivity of devices have resulted in the availability of vast amounts of data. This abundance of big data serves as a valuable resource for AI systems, allowing them to continuously learn and improve their performance over time.
- **Advancements in computing power:** the development of more powerful and cost-effective computing technologies, such as cloud computing and graphic processing units (GPUs), has provided the necessary computational resources to train and run complex AI models efficiently. This has significantly accelerated the development and deployment of AI systems.

**Figure 2:** Data availability and computing power are the foundations of AI

Image created by the author with dall-e (openai.com/research/dall-e) – left – and Imgcreator (imgcreator.zmo.ai) – right. Images are the response of each app to a request to produce an image that shows that "data availability and computing power are the foundations of Artificial Intelligence".

In conclusion, the convergence of these factors has created the ideal conditions for businesses, organizations, and individuals across various sectors and industries to increasingly adopt and apply AI to a wider range of activities.

## 2.2 What are the main benefits of using AI?

**AIntern responded that:** although the main benefits of using AI are different depending on the application and industry, AI adoption offers businesses a wide range of benefits across four broad categories:

1. **Operational efficiency:** AI improves operational efficiency by automating repetitive tasks, optimizing resource allocation, and streamlining processes. This leads to time and cost savings, enhanced accuracy, and improved quality in data processing and analysis. Additionally, AI systems can scale seamlessly to handle large volumes of data and workloads, ensuring scalability and adaptability to changing business needs. AI also enhances security measures by detecting and mitigating potential vulnerabilities, protecting sensitive data, and ensuring compliance with security protocols.

2. **Enhanced customer experience:** AI enables personalized and tailored customer experiences through chatbots, virtual assistants, and recommendation systems. By leveraging AI algorithms, businesses can provide accurate and relevant information to customers, enhancing the accuracy and quality of interactions. This leads to improved customer satisfaction, engagement, and loyalty.

3. **Risk management and compliance:** AI helps mitigate risks by detecting fraud, identifying anomalies, and ensuring compliance with regulatory requirements. Machine learning algorithms enhance accuracy in risk assessment and management, enabling organizations to make informed decisions and identify potential risks. As this can be done in real-time and proactively, companies are better equipped to detect and prevent cybersecurity threats and protect sensitive information.

4. **Innovation and competitive advantage:** AI fuels innovation by providing data-driven insights, accelerating product development, and enabling businesses to identify new opportunities and, therefore, stay ahead of the competition.

Firms have already realized these benefits and are adopting AI technologies in many activities. Consequently, this has brought about a paradigm shift in many industries.

## 2.3 How are banks adopting AI?

**AIntern's response was that:** the banking sector is also part of this trend and is adopting AI through a combination of in-house development, partnerships with technology vendors, and collaborations with fintech startups. In order to effectively leverage the potential of AI, banks are investing in AI talent, infrastructure, and data management capabilities.

As not all banks pursue the same goals and have the same priorities, each bank is adopting AI differently. Nevertheless, there are some areas where most banks are adopting AI.

**Fraud detection** is probably the activity where AI is most pervasively used by banks and one of the areas where they have been using this technology for decades. Indeed, AI has already proved to be a valuable tool in this domain. AI algorithms can analyze vast amounts of data in real-time, identifying patterns and anomalies that indicate fraudulent activities. By continuously monitoring transactions and user behavior, AI-powered systems can detect and prevent fraudulent activities, protecting both the bank and its customers. This not only helps in reducing financial losses but also strengthens the trust and confidence of customers in the banking system.

Further to fighting against fraud, banks are using AI also for **risk management and compliance**. AI technologies help automate anti-money laundering (AML) and know your customer (KYC) processes, streamline and enhance regulatory compliance, improve risk assessment accuracy, and strengthen cybersecurity measures, among others. Machine learning algorithms analyze data to identify patterns, detect anomalies, and provide real-time insights for risk mitigation.

All the above helps banks to **improve efficiency and reduce costs**. Moreover, AI helps banks automate various processes such as back-office operations, including data entry, document processing, and record keeping, which were previously time-consuming and prone to errors. This not only improves efficiency and reduces costs in the long run, but it also allows employees to focus on higher-value activities.

With regards to the commercial activities of banks, AI allows banks to improve conversion rates and customer satisfaction. Thanks to the ability of AI to analyze customer data, market trends, and financial indicators, banks can extract valuable insights for better decision making. Thus, AI helps **identify business opportunities, optimize marketing strategies, develop personalized products, predict customer behavior, and offer personalized recommendations**. Moreover, **chatbots** powered by AI allow banks to provide round-the-clock customer support, answer frequently asked questions, and assist with basic banking transactions.

Finally, **AI is also revolutionizing the main banking activity of credit**. Thanks to AI, banks are able to enhance credit scoring and decision-making processes. Sophisticated algorithms allow banks to analyze vast amounts of data, including both traditional credit information and alternative data sources. This enables banks to assess creditworthiness more accurately, make faster and more informed lending decisions, expand access to credit for individuals with limited credit history, and reduce the risk of defaults.

Moreover, AI plays a crucial role in automating credit underwriting processes, improving risk assessment models, and optimizing loan pricing. Additionally, AI-powered systems have the capability to monitor and detect early signs of credit delinquency, empowering banks to take proactive measures and mitigate potential risks.

### 2.4 Which ethical, privacy, and societal considerations does AI raise for banks?

**AIntern's response was that:** it's important to note that while AI offers numerous benefits, it also raises ethical, privacy, and societal considerations that need to be addressed for responsible and sustainable AI implementation by banks. These considerations include:

1. **Privacy and data protection:** banks handle sensitive customer data, and the use of AI raises concerns regarding data privacy and protection. It is crucial for banks to ensure that AI systems comply with relevant data protection regulations and implement robust privacy measures to safeguard customer information.

2. **Bias and fairness:** AI algorithms can inadvertently reflect biases present in the data used to train them, leading to discriminatory outcomes that in the most extreme case could have wider societal implications, including exacerbating existing social inequalities. Banks must, therefore, ensure that AI systems are developed and trained in a way that minimizes biases and promotes fairness in decision making, particularly in areas such as loan approvals and credit scoring.

**Figure 3:** How banks are using AI


Image created by the author with Imgcreator (imgcreator.zmo.ai) and MS PowerPoint.
Images are responses of Imgcreator (imgcreator.zmo.ai) to request of creating images on the topics indicated.

**Figure 4:** AI's ethical, privacy, and societal considerations


Image created by the author with Imgcreator (imgcreator.zmo.ai) and MS PowerPoint.

3. **Transparency and explainability:** AI models often operate as black boxes, making it difficult to understand the reasoning behind their decisions. Banks need to prioritize transparency and explainability, ensuring that AI systems provide clear explanations for their outputs, especially when they impact customer interactions or financial decisions.

4. **Accountability and liability:** as AI systems make autonomous decisions, determining accountability and liability becomes complex. Banks must establish mechanisms to assign responsibility for AI-related actions and ensure appropriate governance and oversight frameworks are in place.

5. **Customer trust and engagement:** the use of AI in customer interactions, such as chatbots and virtual assistants, raises concerns about customer trust and engagement. Banks need to strike a balance between automated AI interactions and maintaining personalized and human-centric customer experiences.

6. **Ethical use of customer data:** AI applications rely on vast amounts of customer data. Banks must use data ethically, ensuring they have an appropriate basis for processing those data, and using it solely for legitimate purposes while maintaining transparency about data collection and usage practices.

7. **Robust security and safety measures:** AI systems can be vulnerable to attacks and manipulation, posing risks to the security and safety of banking operations and customer data. Banks should invest in robust cybersecurity measures to protect against AI-related threats and ensure system integrity.

Addressing these ethical, privacy, and societal considerations requires banks to establish clear policies and risk mitigation frameworks that ensure the responsible and ethical use of AI technology.

Indeed, regulation can help banks to develop policies and frameworks that ensure responsible AI adoption, customer protection and risk management and, at the same time, maintains financial system stability and integrity.

Nevertheless, it is important to strike a balance in the regulation of AI, ensuring that it addresses potential risks and ethical concerns while allowing for innovation, competition, and the realization of AI's potential benefits.

### 2.5 What can regulators do to support the adoption of AI by banks?

**AIntern's response was that:** a flexible and adaptive regulatory approach that promotes responsible AI development and use is crucial for fostering a thriving and sustainable AI ecosystem. In order to realize this approach regulators shall:

- Provide **clear guidelines and standards** that outline the expectations for AI adoption in the banking sector. This would help banks understand the compliance requirements and facilitate the implementation of AI systems.
- **Collaborate and actively engage** with banks and other stakeholders in the development of AI regulations and policies. Collaboration between regulators and banks allows for a better understanding of the opportunities and challenges associated with AI adoption. Regular dialogue and consultation can help regulators stay informed about technological advancements and evolving industry practices.
- **Support experimentation and innovation** through the creation of sandboxes or innovation hubs where banks can test and experiment with AI technologies in a controlled environment.
- Adopt a **flexible and proportionate approach** towards AI adoption. They can tailor regulatory requirements based on the complexity and risk profile of each potential use case.
- **Collaborate with their international counterparts** to promote consistency and harmonization in AI regulations. This collaboration will help to avoid fragmentation and conflicting regulations, facilitating cross-border adoption of AI technologies by banks. Regulatory cooperation can also enable the sharing of best practices and experiences, fostering a global regulatory environment supportive of AI adoption.
- **Continuously monitor** the use of AI in banks and evaluate its impact on the financial system. This monitoring helps identify emerging risks, potential biases, and ethical concerns associated with AI adoption. Regulators can then update guidelines and regulations to address these issues, ensuring that AI adoption remains aligned with regulatory objectives.

By providing clear guidance, fostering innovation, and maintaining an open dialogue with banks, regulators can support the responsible adoption of AI in the banking sector, promoting efficiency, competitiveness, and customer protection.

### 2.6 What do banks need to do to adopt AI successfully?

**AIntern's response was that:** to adopt AI successfully, banks need to take the following steps:

1. **Develop a clear strategy:** banks should define their goals and objectives for adopting AI and align them with their overall business strategy. This includes identifying areas where AI can add value, such as customer service, risk management, fraud detection, or process automation.

2. **Invest in data infrastructure:** AI relies on large amounts of high-quality data. Banks should invest in data infrastructure to collect, store, and process data effectively. This may involve implementing data management systems, data governance frameworks, and data integration capabilities.

3. **Build internal expertise:** banks need to build a team of AI experts, including data scientists, machine learning engineers, and AI researchers. They can hire new talent or upskill existing employees through training programs. This internal expertise is crucial for developing and deploying AI models effectively.

4. **Collaborate with external partners:** banks can collaborate with external partners, such as fintech companies or technology providers, to access specialized AI capabilities or innovative solutions. Partnerships can help banks accelerate their AI adoption and stay ahead of competitors in the rapidly evolving AI landscape.

5. **Ensure regulatory compliance and ethics:** banks need to comply with relevant regulations and industry standards when implementing AI. They should also consider ethical implications and ensure an appropriate level of transparency, fairness, and accountability in their AI systems.

6. **Start with pilot projects:** banks can start with small-scale pilot projects to test and validate AI solutions before scaling them up. This iterative approach allows banks to learn from initial deployments, make improvements, and mitigate risks.

7. **Monitor and evaluate performance:** banks should continuously monitor and evaluate the performance of their AI systems. This involves measuring key metrics, analyzing outcomes, and making adjustments as needed. Ongoing evaluation helps banks optimize their AI models and ensure they deliver the expected benefits.

8. **Foster a culture of innovation:** banks should promote a culture of innovation and encourage experimentation with AI. This involves creating an environment where employees feel empowered to explore new ideas, take calculated risks, and learn from failures. Innovation should be ingrained in the bank's DNA.

By following these steps, banks can adopt AI successfully and leverage its capabilities to enhance customer experiences, optimize operations, and drive competitive advantage in the digital banking landscape.

**Figure 5:** How banks can successfully adopt AI

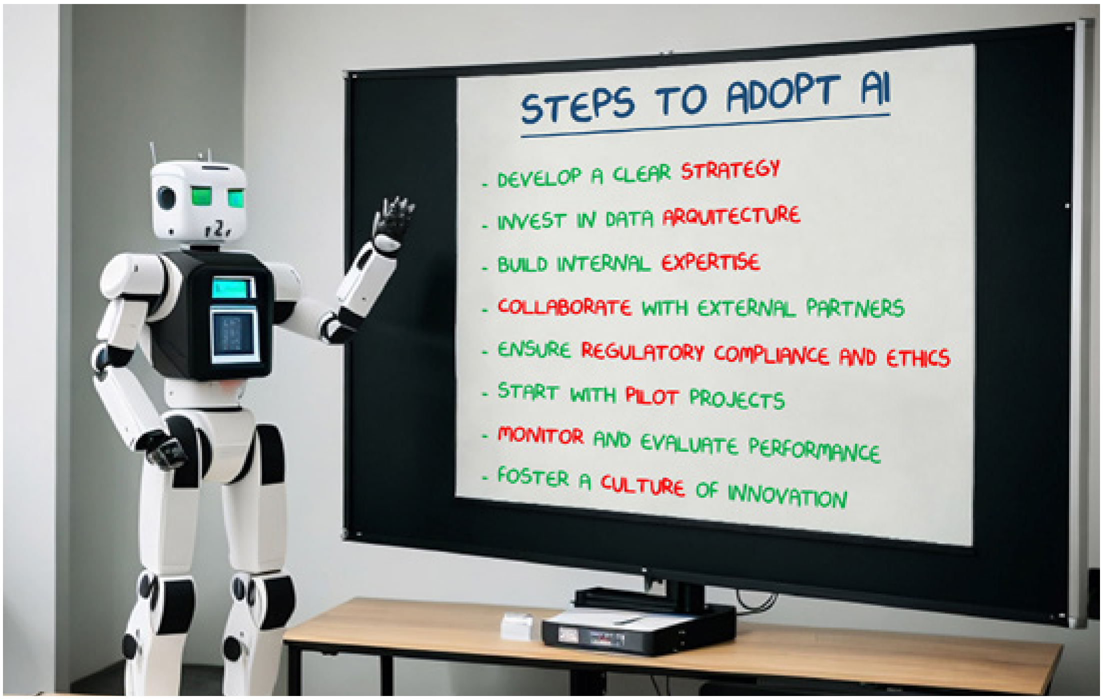



Image created by the author with Imgcreator (imgcreator.zmo.ai) and MS PowerPoint.

## 3. CONCLUSION

The rapid advancements in technology witnessed in recent years, coupled with the increased availability of data and computing power, have facilitated the rapid development of AI. This transformative technology is revolutionizing various industries, including banking, by automating complex tasks such as research and content generation, as exemplified in this article.

The adoption of AI by companies is already yielding numerous benefits, including improved operational efficiency, enhanced customer experiences, and more effective risk management, compliance, and innovation.

In line with this paradigm shift, banks are increasingly adding AI to various processes and services. However, it is crucial for banks to acknowledge the limitations of AI and proactively address the ethical and legal considerations associated with its implementation.

The approach of authorities to AI is essential in shaping those legal considerations. In addition to establishing legal safeguards and potentially prohibiting high-risk applications, authorities should actively engage with companies and organizations on their AI journey. This engagement should involve the issuance of clear guidelines and standards, the provision of a safe environment for AI experimentation, and international collaboration.

Should such a clear and flexible regulatory framework be provided, banks will be able to design medium and long-term strategies that encompass investments in infrastructure and human capital, as well as a gradual and responsible adoption of AI.

## APPENDIX: HOW I WROTE THIS ARTICLE

| NR | STEP | OBSERVATIONS |
|---|---|---|
| 1 | Choose the topic of the article:<br>"The use of AI in banks. Benefits, challenges, and regulatory considerations" | |
| 2 | Choose the questions to be asked:<br>1. Why is this the moment for AI?<br>2. What is needed to extract value from data and/or AI?<br>3. How can banks take advantage of AI?<br>4. Which banking activities would benefit more from the application of AI?<br>5. How are banks adopting AI?<br>6. What is the role of Regulation in the bank's adoption of AI?<br>7. What do banks need to do to adopt AI successfully?<br>8. What can regulators do to support the adoption of AI by banks? | |
| 3 | Select reports, papers, and articles:<br>I chose reports, news, and papers I have written or contributed to and other documents I have read since 2017 that I found interesting, and all of which are publicly available | Need to take into consideration copyright of material to be uploaded and the terms and conditions of the AI tools to be used, since they can store and use the documents uploaded for purposes banned by the copyright of the material. |
| 4 | Search for additional sources:<br>elicit.com and you.com | Although the average quality of papers that can be found in elicit is arguably higher, most of the material located through elicit is subject to distribution and usage restrictions. |
| 5 | Upload sources (in pdf format) in AI tool:<br>ayfie.com | I located ayfie in humanalternative.com, a site that curates AI tools to automate several different tasks. I chose ayfie.com because it allows users to upload up to 1,000 files and query all files at a time or a group of them. |
| 6 | Ask questions listed in 2 through the chat functionality | When the answer did not meet my expectations (because it was too long or too short, or it missed a piece of information that I considered relevant), I refined the question or added additional conditions and submitted it again. |
| 7 | Copy and edit satisfactory answers in a separated document | The answers provided are not attributed to specific sources, so it is not clear how they have been produced. It is important to review the text to remove erroneous and redundant information. Be aware of hallucinations of facts, especially in topics you are not an expert in. |
| 8 | Merge answers and improve drafting if necessary:<br>Chatsonic functionality in writesonic.com | Answers tended to be lists of items that reiterated some ideas. |
| 9 | Review and edit the text again and add new content, if necessary | The author must feel comfortable with the final text. |
| 10 | Add charts and images that support text.<br>www.aipicasso.app, imgcreator.zmo.ai, and openai.com/research/dall-e | Getting the correct images and charts requires practice and accuracy. I used some output images as input for other tools and I retouched the final output with traditional software. |

## REFERENCES

Acemoglu, D., 2021, "Harms of AI," NBER working papers 29247, National Bureau of Economic Research

Alonso, A., and J. M. Carbó, 2022, "Inteligencia artificial Y finanzas: Una Alianza estratégica (artificial intelligence and finance: A strategic alliance)," SSRN

AFME, 2018a, "Artificial intelligence: adoption in capital markets," Association for Financial Markets in Europe

AFME, 2018b, "Considerations on the ethical use of artificial intelligence in capital markets," Association for Financial Markets in Europe

AFME, 2019, "Artificial intelligence and machine learning in capital markets: considerations for a broad framework for transparency," Association for Financial Markets in Europe

Benaich, N., and I. Hogarth, 2022, "State of AI report 2022," https://tinyurl.com/yr6neds3

Boukherouaa, E. B., M. G. Shabsigh, K. Aiajmi, J. Deodoro, A. Farias, E. S. Iskender, M. A. T. Mirestean, and R. Ravikumar, 2021, "Powering the digital economy: opportunities and risks of artificial intelligence in finance," International Monetary Fund, https://tinyurl.com/3z84jkxw

Bughin, J., E. Hazan, S. Ramaswamy, M. CHui, T. Allas, P. Dahlström, N. Henke, and M. Trench, 2017, "Artificial Intelligence. The next digital frontier?" McKinsey Global Institute, https://tinyurl.com/2kaces57

Chui, M., E. Hazan, R. Roberts, A. Singla, K. Smaje, A. Sukharevsky, L. Yee, and R. Zemmel, 2023, "The economic potential of generative AI: the next productivity frontier," McKinsey & Company

Clifford Chance and Milltown Partners, 2021, "Our relationship with Ai: friend or foe a global study," https://tinyurl.com/y7xnyvfd

Deloitte AI Institute, 2023, "A new frontier in Artificial Intelligence. Implications of Generative AI for businesses," Deloitte

Deloitte University, 2014, "Demystifying artificial intelligence," Deloitte

EBA, 2020, "Final report on big data and advanced analytics," European Banking Authority

EBF, 2020, "AI in the banking industry: EBF position paper," European Banking Federation

Fernández, A., 2019, "Artificial intelligence in financial services," Economic Bulletin 2/2019

FSB, 2017, "Artificial intelligence and machine learning in financial services," Financial Stability Board

Fintech Global, 2023, "AIFinTech100," https://tinyurl.com/23z4av54

Gerbert, P., and M. Hecker, 2017, "Competing in the age of artificial intelligence," BCG Global, Boston Consulting Group

Ghandour, A., 2021, "Opportunities and challenges of Artificial Intelligence in banking: systematic literature review," TEM Journal, 1581–1587.

IIF, 2019a, "Bias and ethical implications in machine learning," Institute of International Finance

IIF, 2019b, "Machine learning recommendations for policymakers," Institute of International Finance

IIF, 2020, "Machine learning governance summary report," Institute of International Finance

IIF, 2022, "IIF and EY 2022 Survey Report on machine learning – summary," Institute of International Finance

Lozano, J., 2018, "Hacia una inteligencia artificial controlada y ubicua," Ediciones EL PAÍS S.L.

Lozano, J., 2022, "Cómo se debe regular la inteligencia artificial? BBVA, https://tinyurl.com/2brzw328

Lozano, J., and J. Murillo, 2020, "Qué habría que tener en cuenta si se va a regular la inteligencia artificial?" BBVA

Lozano, J., T. Rodrigo, and J. Sicilia, 2020, "Fronteras del big data para el análisis económico y financiero en banca," Papeles de Economía Española 162, 98–116

Maslej, N., L. Fattorini, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, H. Ngo, J. C. Niebles, V. Parli, Y. Shoham, R. Wald, J. Clark, and A. R. Perrault, 2023, "The AI Index 2023 annual report," https://tinyurl.com/mcrc2tep

McCauley, D., 2016, "Ghosts in the machine. Artificial intelligence, risks and regulation in financial markets," Baker & McKenzie and Euromoney.

McKinsey & Co., 2023, "What is generative AI?" https://tinyurl.com/5dxbk4da

NVIDIA, 2021, "The state of Ai in financial services," https://tinyurl.com/5e5af24r

OECD, 2020, "Identifying and measuring developments in artificial intelligence: making the impossible possible," https://tinyurl.com/mr27y8t3

OECD, 2021, "Artificial Intelligence, machine learning and big data in finance: opportunities, challenges, and implications for policy makers," https://tinyurl.com/2y3usn6m

Renda, A., 2019, "Artificial Intelligence: ethics, governance and policy challenges," SSRN

Stancombe, C., R. Tolido, J. Buvat, A.-L. Thieullent, A. Khadikar, K. V. J. Subrahmanyam, and A. Chandna, 2018, "Turning AI into concrete value. The successful implementers toolkit," CapGemini, https://tinyurl.com/5y25nc5m

Tobin, J., 2023, "Artificial Intelligence: development, risks and regulation," U.K. Parliament, https://tinyurl.com/mrx268vn

WEF, 2019, "AI governance: a holistic approach to implement ethics into AI," World Economic Forum, https://tinyurl.com/yckb5e93

# PERFORMANCE OF USING MACHINE LEARNING APPROACHES FOR CREDIT RATING PREDICTION: RANDOM FOREST AND BOOSTING ALGORITHMS

**W. PAUL CHIOU** | Associate Teaching Professor of Finance, Northeastern University
**YUCHEN DONG** | Senior Engineer, MathWorks
**SOFIA X. MA** | Senior Engineer, MathWorks

## ABSTRACT

Applying machine learning techniques to improve the accuracy and efficiency of predictions of credit risk rating is increasingly critical to the financial services industry. In this study, we apply MATLAB to investigate the performance of two approaches, decision forest and boosting algorithms, by using a wide range of financial data. The empirical outcomes suggest that both methods exhibit considerable performance but may be superior to each other in different scenarios. Boosting algorithms method exhibits an accuracy rate of approximately 67% across the credit rating categories. The random forests model generates lower accuracy rates for low and medium classifications than the boosting method, but the accuracy rate for high credit ratings reaches 79%, more accurate than the boosting method.

## 1. INTRODUCTION

Credit rating prediction is a critical task in the financial services industry, as the outcomes can affect investment decisions, corporate finance, and risk management. An accurate forecast of default risk provides an early warning system for identifying entities or investments that may pose financial, operational, or strategic risks. With recent systematic shocks, such as the COVID-19 pandemic, risk assessment and compliance required by regulators make predicting credit rating essential to avoid legal repercussions. As financial institutions use risk ratings to make lending decisions and to determine interest rates, accurate risk ratings can be critical to managing corporate finance. However, credit risk is modeled under assumptions of trackable borrower and market dynamics and does not account for unforeseen events, hence, leaving the models unable to produce reliable results.

Machine learning techniques have been widely applied for their potential to improve the accuracy and efficiency of predictions. However, several challenges – such as how to process data to ensure the quality for analysis, imbalanced numbers of defaults compared to non-defaults, identifying the relevant features, model validation to ensure accuracy and robustness, and selecting the appropriate algorithm – still need to be addressed in applying machine learning to forecast credit risk ratings.

This research contributes to the literature threefold. First, the levels of accuracy and interpretability of credit risk predictions may vary across different algorithms; consequently, identification of the appropriate approaches remains uncertain. Second, validation processes that consider performance metrics and testing scenarios can be complex and are relevant to the usefulness of the selected models. Third, identifying the relevant credit-related factors that can help minimize the impacts of feature overload, the curse of dimensionality, multicollinearity, and noise in data is crucial to risk management practice.

This study applies two methods, decision tree boosting and bagging algorithms, and highlights their relative strengths and applicability to credit rating prediction using a wide range of financial data as input features. Machine learning techniques have advanced credit risk assessment but can be highly complex, leaving implementation and interpretation of the outcomes challenging. We, particularly, apply the machine learning tools in MATLAB, specifically, bagged random forests and boosting algorithms, which are used in credit rating predictions. The findings suggest that both methods exhibit considerable performance but may be superior to one another in different scenarios. Boosting algorithms method exhibits accuracy rates of approximately 67% across the credit rating categories. The random forest model generates lower accuracy rates for low and medium classifications than the boosting method, but the accuracy rate for high credit ratings reaches 79%, more accurate than the outcome using the boosting method.

## 2. MODELS

### 2.1 Random forests

Random forests, or decision forests, that assemble a collection of decision trees working conjointly in predictions and classifications belong to a family of supervised machine learning models and algorithms. The methods offer numerous advantages, such as ease of configuration, native handling of diverse features, robustness to noisy data, and interpretability. Due to properties such as interpretability, scalability, resistance to overfitting, and handling missing data, decision forests are suitable for signal integration from tabular data, allowing for efficient aggregation of signals from multiple subsystems.

Random forests serve as a remedy for the overfitting issues related to the tree learning approach, such as low bias but high variance, resulting in decreased accuracy. Working on decision forests involves creating and training multiple decision trees with random subsets of data and features. By combining multiple deep decision trees trained on different subsets of the training data, this approach reduces variance while introducing a slight increase in bias and some loss of interpretability. The teamwork of many trees in a forest effectively enhances the overall performance, as compared to a single random tree, yielding more accurate and robust results for data mining tasks.

To build decision forests, the bagging method, repeatedly selecting random samples with replacements from the training set, is utilized in the training algorithm. Classification or regression trees are trained using these subsamples, and they predict unseen samples by either averaging the predictions from all individual regression trees or taking the majority votes from classification trees. The above procedure improves model performance by reducing variance while controlling the increase in bias. By creating fewer correlated trees through different training sets, bagging ensures that the average predictions of multiple trees are less sensitive to noise, as compared to a single tree. The process involves selecting a random subset of features at each candidate split, mitigating the issue of strong predictors dominating multiple trees and causing correlation.

Random forests are trained as a system that has few hyperparameters that can be of proper default values. This allows them to be more efficient in data preprocessing while reducing error sources and enhancing the accuracy of the results. A group of decision trees utilizes a random subset of features and data points from the training set, allowing numeric features to be natively handled. This enables the generation of robust results from highly stochastic data. Thus, the results of decision forests can be easily interpretable and understood.

As a supervised learning model, bagging is usually used to reduce the variance of the decision trees by averaging the prediction over a collection of bootstrap samples. Specifically, bagging is to create several subsets of data from the training samples chosen randomly with replacement. As a result, the prediction from sampled data will be more robust than using only one single decision tree. Suppose the training data is $\mathbf{Z} = \{(x_1, y_1), \ldots(x_N, y_N)\}$. Our goal is to determine the prediction $\hat{f}(x)$ given the data x. Denote the bootstrap samples as $\mathbf{Z}^b$, where $b = 1, 2, \ldots, B$. Here, $B$ is the number of bootstrap sampled dataset. The bagging estimate is defined as

$$f_{\text{bag}}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x) \quad (1)$$

Considering equation (1) as the Monte Carlo estimation, it converges to the true estimation as $B$ goes to infinity. Let us consider the regression tree model as an example. Let $\hat{f}^{(x)}$ be the tree's prediction given the input data $x$. The trained tree model from bootstrapping samples typically involves different features than the original one. It might have a different number of terminal nodes as well. The bagged estimate $\hat{f}_{\text{bag}}(x)$ is the average prediction from $B$ different trees.

In addition to the regression tree, the classification tree model is also popular and used in the following example. Suppose the target of the classification model takes value in 1, 2, ..., $K$, and we have $m$ nodes and region $R_m$ with $N_m$ observations. We further define the proportion of class $k$ observations in node $m$ as

$$\hat{p}_{m,k} = \frac{1}{N_m} \sum_{x_i \in R_m} 1_{\{y_i=k\}} \tag{2}$$

We classify the observations in node $m$ to class $k(m) = argmax_K\, \hat{p}_{m,k}$, the majority class in node $m$. Considering the cost complexity criterion, we can select different measures of node impurity, such as misclassification error, Gini index, or cross-entropy based on different situations.

## 2.2 Boosting algorithms

Boosting algorithms is an ensemble learning technique that combines multiple weak decision tree models to enhance predictiveness through iterative learning steps. It is an extension of the boosting method, which is a general approach for improving the performance of a base learning algorithm by combining several weak learners in a weighted manner.

The boosting algorithm offers an effective solution for prediction tasks in both classification and regression fields. By iteratively adding decision trees to the ensemble, the framework allows each new tree to be trained to correct the errors of its predecessor. The trees are added sequentially, with each tree learning to fit the residual errors from the previous trees. In each iteration, coefficients, weights, or biases of input variables are adjusted to minimize the loss function, measuring the discrepancy between predicted and actual target values. The gradient represents incremental adjustments in each step, while boosting accelerates predictive accuracy improvements, reaching an optimal level. The final model is the weighted sum of all individual trees. By streamlining the objective and minimizing iterations, this method enhances the learning process, achieving a satisfactory optimal solution more efficiently.

The machine learning tools in MATLAB 2023a, such as bagged decision trees, are used in the domain of credit rating prediction in this study. The flexibility of the software, such as Deep Learning Toolbox and Database Toolbox, enables researchers to tailor and adapt the workflow delineated to their unique preferences and specific requirements.

# 3. DATA, PROCESS, AND PREPARATION FOR ANALYSIS

Financial ratios as predictors are used to forecast the credit rating as the response variable by fitting a bagged decision tree. Bagging, or bootstrap aggregation, consists of generating many random sub-samples, or bootstrap replicas from the dataset by sampling with replacement from the list of customers in the dataset. A decision tree grows from the replica. Each decision tree is a trained classifier on its own and could be used in isolation to classify new clients. The predictions of two trees grown from two different bootstrap replicas may be different. The ensemble aggregates the predictions of all the decision trees that are grown for all the bootstrap replicas.

**Table 1:** Variables used to predict credit rating

| RATIO | MEDIAN | RATIO | MEDIAN |
|---|---|---|---|
| Current ratio | 1.49 | Debt equity ratio | 1.65 |
| Quick ratio | 0.99 | Debt ratio | 0.64 |
| Cash ratio | 0.30 | Effective tax rate | 0.30 |
| Days of sales outstanding | 42.37 | Free cash flow operating cash flow ratio | 0.64 |
| Net profit margin | 0.06 | Free cash flow per share | 2.13 |
| Pretax profit margin | 0.08 | Cash per share | 3.69 |
| Gross profit margin | 0.41 | Company equity multiplier | 2.65 |
| Operating profit margin | 0.11 | EBIT per revenue | 0.09 |
| Return on assets | 0.05 | Enterprise value multiple | 9.27 |
| Return on capital employed | 0.07 | Operating cash flow per share | 4.35 |
| Return on equity | 0.12 | Operating cash flow sales ratio | 0.13 |
| Asset turnover | 0.70 | Payables turnover | 5.76 |
| Fixed asset turnover | 3.81 | | |

If the majority of the trees predict one particular class for a new customer, it is reasonable to consider that prediction to be more robust than the prediction of any single tree alone. The information is still useful when a different class is predicted by a smaller set of trees. The proportion of trees that predict different classes is the basis for the classification scores that are reported by the ensemble when classifying new data.

### 3.1. Data

The quarterly data of 2,029 credit ratings between 2010 and 2016 are used in this study. Table 1 presents the financial ratios applied in this study to forecast credit rating and their medians. These widely applied measures collectively provide insights into a company's liquidity, profitability, asset management efficiency, and financial leverage, and are essential for assessing financial health, and risk profile, widely regarded as possible factors for predicting credit ratings, such as Altman's z-score (1968). First, it is natural to consider financial leverage that involves the use of debt to finance a company's operations. The debt/equity ratio of 1.65 and debt ratio of 0.64 represent, overall, the reliance on debt of companies in their capital structures. A higher debt/equity ratio indicates a relatively higher level of debt compared to equity, while the debt ratio illustrates the proportion of total assets financed by debt.

To measure the ability to meet short-term obligations, current ratio, quick ratio, and cash ratio provide insights into a company's liquidity position. Several ratios are used to measure profitability from different aspects: net profit margin, pretax profit margin, gross profit margin, operating profit margin, return on assets, return on capital employed, and return on equity. The fact that the company retained, on average, $0.06 in profit for every $1 in net sales revenue over the sample period, similar to the historical averages, validates the use of the data for credit risk analysis. In addition, utilization of assets is also considered in the analysis as operational performance can be critical to risk management. The "days of sales outstanding" suggests that the companies, overall, take an average of 42.37 days to convert sales into cash receipts. Asset turnover and fixed asset turnover values of 0.70 and 3.81, respectively, indicate the operating efficiency of the overall assets and fixed assets in generating revenue.

The distribution of sectors in this study includes a range of industries. The largest sectors by number of observations are energy, consumer services, public utilities, technology, and basic industries, together representing more than 60% of the sample analyzed. Other than the sectors above, the majority of companies included are from manufacturing industries. On the other hand, the study only includes 50 observations from the financial services sector. As the sample of this study reflects a comprehensive analysis of non-financial service industries, the financial ratios applied will be meaningful to determine the credit risk.

For the distribution of credit ratings, the majority falls within the investment-grade categories of BBB (671) and A (398), while higher credit ratings, such as AAA (7) and AA (89), are less common. Riskier credit ratings, such as CCC-rated or lower, comprise smaller portions of the dataset, representing less than 4% of the sample. The distribution of raters in the study shows that Standard & Poor's has the highest representation, followed by Moody's, and Egan-Jones, representing about 95% of the samples. Other rating agencies, like Fitch and DBRS, have comparatively fewer observations.

**Table 2:** The distribution of companies and rating agencies

| SECTOR | N | RATING | N | RATING AGENCY | N |
|---|---|---|---|---|---|
| Basic industries | 260 | AAA | 7 | DBRS | 3 |
| Capital goods | 234 | AA | 89 | Egan-Jones | 603 |
| Consumer durables | 73 | A | 398 | Fitch | 100 |
| Consumer non-durables | 132 | BBB | 671 | Moody's | 579 |
| Consumer services | 250 | BB | 490 | Standard & Poor's | 744 |
| Energy | 294 | B | 302 | | |
| Finance | 50 | CCC | 64 | | |
| Healthcare | 171 | CC | 5 | | |
| Public utilities | 211 | C | 2 | | |
| Technology | 234 | D | 1 | | |
| Transportation | 63 | | | | |
| Miscellaneous | 57 | | | | |

### 3.2. Characteristics across various ratings

Table 3 presents a summary of various financial indicators across different credit ratings, providing a concise overview of key financial and operational indicators across different credit ratings and the first look at the relationships between creditworthiness and various performance metrics. The financial ratios in the empirical analysis include various measures of liquidity, profitability, asset management, and financial leverage. Companies with higher credit ratings tend to exhibit more favorable financial metrics. As shown in Panel A, higher credit ratings, such as AAA and AA, are associated with stronger liquidity, as seen in their higher current ratios and quick ratios compared to lower-rated categories. For instance, companies rated AAA showcase a high current ratio (CR) of 2.50, indicating a strong ability to cover short-term liabilities. Days of sales outstanding tend to decrease as credit ratings improve, indicating better management of receivables. Operating cash flow, generally, increases with higher credit ratings, except for a dip in the CCC category. Free cash flow per share and cash per share also tend to be more favorable in higher credit rating tiers, with the AAA-rated showing the strongest positions and the D-rated exhibiting the weakest.

**Table 3:** Summary statistics of some variables

| PANEL A: LIQUIDITY | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| FINANCIAL RATIO/RATING | AAA | AA | A | BBB | BB | B | CCC | CC | C | D |
| Current ratio | 2.50 | 1.47 | 1.34 | 1.43 | 1.67 | 1.62 | 1.68 | 1.34 | 1.52 | 0.59 |
| Quick ratio | 2.30 | 0.97 | 0.85 | 0.93 | 1.14 | 1.06 | 1.17 | 0.57 | 0.55 | 0.39 |
| Cash ratio | 0.19 | 0.34 | 0.25 | 0.28 | 0.37 | 0.29 | 0.37 | 0.30 | 0.37 | 0.02 |
| Days of sales outstanding | 78.13 | 39.19 | 39.96 | 42.37 | 43.88 | 43.98 | 41.72 | 21.11 | 14.20 | 54.74 |
| Operating cash flow per share | 3.88 | 8.31 | 5.57 | 5.07 | 3.58 | 2.14 | 1.44 | -1.51 | 4.54 | 3.22 |
| Operating cash flow sales | 0.37 | 0.18 | 0.15 | 0.15 | 0.12 | 0.10 | 0.06 | 0.00 | 0.03 | 0.08 |
| Free cash flow operating cash flow | 0.80 | 0.71 | 0.69 | 0.60 | 0.66 | 0.63 | 0.78 | 1.00 | -0.33 | 0.42 |
| Free cash flow per share ($) | 3.22 | 3.61 | 3.20 | 2.42 | 1.65 | 0.75 | 0.01 | 1.75 | 1.94 | 1.36 |
| Cash per share ($) | 10.33 | 6.07 | 3.91 | 3.65 | 3.78 | 2.26 | 3.79 | 7.75 | 12.19 | 1.26 |
| PANEL B: PROFITABILITY | | | | | | | | | | |
| Net profit margin | 0.20 | 0.11 | 0.09 | 0.07 | 0.05 | 0.02 | -0.03 | -0.23 | -0.30 | 0.18 |
| Pretax profit margin | 0.08 | 0.15 | 0.12 | 0.09 | 0.06 | 0.02 | -0.04 | -0.10 | 0.03 | 0.13 |
| Gross profit margin | 0.69 | 0.59 | 0.49 | 0.38 | 0.35 | 0.41 | 0.76 | 0.92 | 0.11 | 1.00 |
| Operating profit margin | 0.30 | 0.15 | 0.14 | 0.12 | 0.08 | 0.06 | 0.02 | -0.09 | -0.01 | -0.20 |
| Return on assets | 0.10 | 0.09 | 0.07 | 0.05 | 0.03 | 0.02 | -0.02 | -0.33 | -0.34 | 0.04 |
| Return on capital employed | 0.06 | 0.17 | 0.13 | 0.08 | 0.05 | 0.03 | -0.01 | -0.61 | -0.02 | 0.04 |
| Return on equity | 0.23 | 0.18 | 0.17 | 0.13 | 0.09 | 0.06 | -0.01 | 0.91 | 0.64 | -0.50 |
| EBIT per revenue | 0.08 | 0.15 | 0.13 | 0.10 | 0.06 | 0.03 | -0.04 | -0.10 | 0.03 | 0.13 |
| PANEL C: ASSET MANAGEMENT | | | | | | | | | | |
| Asset turnover | 0.53 | 0.80 | 0.71 | 0.70 | 0.71 | 0.65 | 0.41 | 0.51 | 1.18 | 0.24 |
| Fixed asset turnover | 6.35 | 3.96 | 4.48 | 3.54 | 4.11 | 3.38 | 2.14 | 3.08 | 4.85 | 0.26 |
| Effective tax rate | 0.28 | 0.29 | 0.31 | 0.31 | 0.30 | 0.26 | 0.17 | 0.00 | -0.48 | -0.39 |
| Payables turnover | 4.19 | 5.48 | 4.99 | 5.76 | 6.46 | 6.08 | 1.84 | 1.12 | 11.82 | 3.65 |
| PANEL D: FINANCIAL LEVERAGE | | | | | | | | | | |
| Debt equity ratio | 0.92 | 1.37 | 1.59 | 1.67 | 1.6 | 2.09 | 2.39 | -3.46 | -2.91 | -12.41 |
| Debt ratio | 0.48 | 0.58 | 0.62 | 0.63 | 0.64 | 0.74 | 0.79 | 1.21 | 1.65 | 1.09 |
| Company equity multiplier | 1.92 | 2.28 | 2.56 | 2.68 | 2.6 | 3.16 | 3.46 | -2.75 | -1.91 | -11.41 |
| Enterprise value multiple | 11.33 | 9.79 | 9.82 | 9.2 | 8.59 | 9.46 | 8.93 | -1.69 | 1.68 | 20.82 |

The findings in Panel B suggest a correlation between the financial health and profitability of companies. Higher credit-rated companies, such as AAA and AA, exhibit stronger profitability metrics, including higher net profit margin, pretax profit margin, gross profit margin, and return on equity. In contrast, lower credit-rated categories, like B, CCC, and below, show weaker profitability indicators, with negative values observed for net profit margin, pretax profit margin, and operating profit margin in some cases. Return on assets and return on capital employed tend to decrease as credit ratings decline, while return on equity displays a mixed trend. Distinct trends of asset management efficiency across different credit ratings are evident in Panel C. The companies with higher credit ratings display more efficient asset utilization, evident from higher asset turnover and fixed asset turnover ratios, while maintaining a moderate effective tax rate. Conversely, companies in lower credit-rated categories, particularly C- and D-rated, exhibit varying asset management efficiency, with lower asset turnover and fixed asset turnover. The payables turnover ratio is less consistent but generally tends to be higher for higher credit-rated categories.

The financial leverage across different credit ratings reveals several patterns. In Panel D, companies with higher credit ratings, like AAA and AA, tend to have lower debt equity ratios, debt ratios, and company equity multipliers, indicating more conservative financial structures. On the other hand, the lower-rated exhibits higher financial leverage. Although the enterprise value multiple varies inconsistently across credit ratings, it appears to be lower for the higher-rated, confirming that higher credit-rated companies tend to have lower financial leverage structures.

From the first glance at the data, it becomes evident that the liability burden increases across lower credit ratings, suggesting a higher proportion of debt in their capital structures. Companies with low ratings exhibit poor profitability and inefficient operations, indicating a possibility of financial distress. Moreover, the cash per share (CPS) varies across the credit rating spectrum, reflecting the liquidity position of each category. Overall, the trends suggest a correlation between credit rating and financial health, underscoring the implicit information of these metrics in assessing a company's risk profile and stability. Thus, it makes sense to apply these indicators to determine the financial health and operational efficiency of companies within each rating tier.

### 3.3. Data preprocessing

The process of data cleaning and selection in the context of machine learning is critical to ensure the effectiveness and reliability of the resulting models. Since not all available data may exhibit large and representative characteristics, data preprocessing stands out as a pivotal stage in the machine learning algorithm. After observing the data, we find that certain credit ratings, such as AAA and D, have insufficient representation. For instance, AAA rating only has seven data points and rating D only has one data point. Thus, data consolidation into fewer categories is essential to address this issue. After the rating is reorganized as demonstrated in Table 4, the risk data is representative, as each classification is large. We further let 80% of data, or 1,623 observations, be tested while the others are used for training.

**Table 4:** Reorganized credit rating

| | CREDIT VALUE | # | % |
|---|---|---|---|
| 1 | Low | 494 | 24.35 |
| 2 | Medium | 671 | 33.07 |
| 3 | High | 864 | 42.58 |

### 3.4. Classifying new data classification

The previously constructed classification can be used for the assignment of credit ratings to new companies. Since the ratings for existing companies also require reviews that account for variations in their financials, the dataset includes a list of such customers. To predict the credit ratings for the new data, the classifier's "predict" method is invoked. This method yields two essential outputs: the predicted class and the associated classification score. Both output arguments are acquired as the classification scores furnish valuable insights into the confidence level associated with the predicted ratings. Certain advanced computational software, such as MATLAB, facilitates an easily applied tool for the report generation of the classification process.

Preserving records of predicted ratings and their corresponding scores can be used to prove the benefits of periodic assessments of classifier performance. This information can be efficiently stored within a table and further archived through means such as saving to a comma-delimited text file or direct integration into a database system.

### 3.5. Back-testing: profiling the classification process

The assessment of model performance and the validation of credit ratings are applied in this research. The first measurement centers on the accuracy of predicted ratings relative to actual ratings. Predicted ratings are derived from automated classification processes, while actual ratings are assigned by a credit committee amalgamating various information sources. The second measurement evaluates the accuracy of actual ratings retrospectively. Specifically, it examines whether the actual ratings effectively mirror the credit risk of customers. The ex-post analysis, generally conducted over a one-year horizon, identifies companies that experienced defaults during the period, assessing actual rating accuracy.

The research leverages ex-post credit rating data, encompassing subsequent developments for the same companies considered previously. This dataset includes ratings assigned by the credit committee, along with a default flag indicating whether a given company defaulted within a year of the rating process.

Enhanced accuracy of predicted ratings translates into enhanced efficiency in reviewing these predictions. Consequently, it is plausible that the credit risk committee seeks periodic evaluations to gauge alignment between predicted and final ratings, potentially recommending re-training of the automated classifier in case of significant disparities. To facilitate the comparison of predicted versus actual ratings, a confusion matrix is employed. The matrix can be normalized by standardizing values as percentages by dividing the number of observations with true ratings.

## 4. EMPIRICAL RESULTS

### 4.1. Boosting algorithms method

We employ the AdaBoostM2 model, a technique for multiclass classification to conduct boosting for the projection of credit risk values (low, medium, high) by using the information of financial ratios. This ensemble method involves the aggregation of multiple weak learners (decision trees), iteratively refining their predictive power. Our dataset comprises 1,623 observations, out of which 150 are used for training. AdaBoostM2 employs a weighted pseudo-loss function to measure classification accuracy, particularly suitable for multiple classes. Initially, the ensemble prioritizes low pseudo-loss values in the early training steps, indicating strong performance from the first few learners. Subsequently, as the ensemble grows, the learning rate diminishes, gradually approaching a pseudo-loss value of 0.5, which represents random chance.

As presented in Figure 1, it is observed that a reduction in error decreases when leaf size increases. Specifically, the impact of altering decision leaf sizes ranging from 1 to 25 can be found as the error drops by 27%, from an initial 0.37 to 0.27. However, the error rate improvement diminishes as the number of trees increases in the ensemble, indicating the diminishing returns of additional trees.

We next evaluate the performance of the model by applying the confusion matrix for each class, expressed as a percentage of the true rating. Specifically, the matrix aims to present the effectiveness of the automated classification process in predicting credit ratings compared to the ratings assigned by human credit rating agents. The first metric is the accuracy of predicted ratings, generated through automated classification, juxtaposed with the actual ratings determined by human agents. These human assessments incorporate a wide array of information, including economic conditions, news, subjective judgment, and potentially other pertinent

**Figure 1:** Train classification error

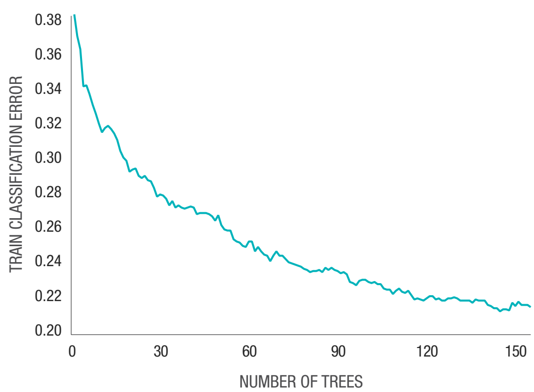



**Figure 2.** Normalized confusion matrix in percentage: boosting method

| TRUE CLASS | 1-low | 2-medium | 3-high | Accuracy | Error |
|---|---|---|---|---|---|
| 1-low | 67.6% | 30.5% | 1.9% | 67.6% | 32.4% |
| 2-medium | 12.5% | 67.2% | 20.3% | 67.2% | 32.8% |
| 3-high | 2.9% | 30.6% | 66.5% | 66.5% | 33.5% |
| | Predicted classes | | | | |

**Figure 3:** Classification error for different leaf sizes

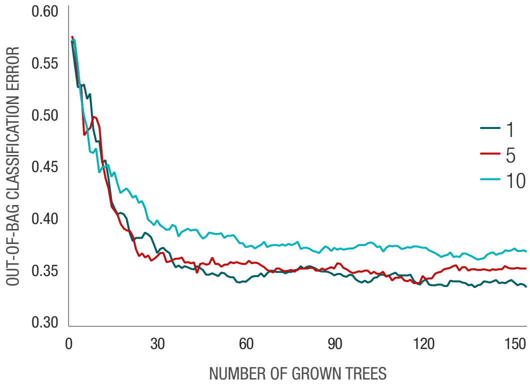
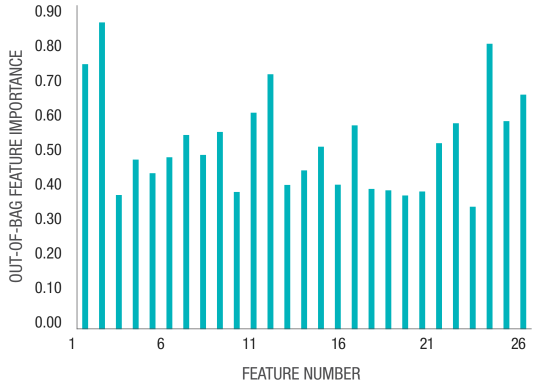



**Figure 4:** Feature importance analysis




data. The second metric pertains to the accuracy of actual ratings that evaluate the extent to which these ratings reflect the default risk of companies. This ex-post analysis involves scrutinizing which companies experienced defaults within a specified period, typically one year. The analysis encompasses follow-up information on previously evaluated companies, encompassing the ratings assigned and a binary flag indicating whether a company defaulted within a year of the rating process.

Since the primary objective of employing an automated classifier is to expedite the work of the credit committee, enhancing accuracy in predicted ratings can improve the efficiency of reviewing these ratings. Consequently, regular assessments are essential to ascertain the alignment between predicted and final ratings. Any significant disparities may trigger recommendations for retraining the automated classifier.

The confusion matrix, illustrated in Figure 2, compares the predicted and actual ratings. We normalize the values by dividing them by the number of observations with the same true rating to ensure a meaningful assessment. An ideal outcome would manifest as values in the main diagonal dominating the other entries in each row, ideally approaching 1. Our model exhibits accuracy rates of approximately 67% across the three credit rating categories.

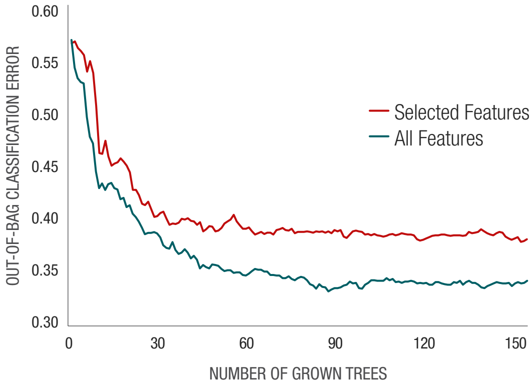
## 4.2. Decision forest method

The first phase of constructing the classification ensemble, or tree bagger, is to determine an optimal leaf size for individual trees. In this setting, there is no requirement to partition the data into distinct training and test subsets, as this partitioning process occurs internally and implicitly during the sampling procedure. The classification error trends for various leaf sizes (1, 5, and 10 in this analysis) are examined while considering a maximum of 150 trees in the ensemble. To ensure reproducibility and facilitate comparisons, the random number generator is reset for each iteration, allowing for the resampling of data for classifier construction. Figure 3 demonstrates the comparable errors observed across the three leaf-size options and suggests that a leaf size of ten is preferable, as it leads to the development of more streamlined trees and enhances computational efficiency.

The training set consists of the bootstrap replica for each bootstrap iteration. For any "out-of-bag" samples, the observations left out are employed as test points to estimate the out-of-bag classification error. To maintain the efficiency of computational processes and yield leaner trees, a sample size of 10 is employed in this study.

The subsequent step involves the assessment of feature importance to discern their contribution to improving the accuracy of the risk classifier. As presented in Figure 4, certain features emerge as pivotal among the feature set. Specifically, the rating agency (Feature 1), industry (Feature 2), return on capital employed (Feature 12), operating cash flow per share (Feature 25), operating cash flow over sales ratio (Feature 26), and payables turnover (Feature 27) stand out as the most influential predictors within this dataset. It is noteworthy that the inclusion of these features underscores their substantial role in predictive accuracy. Furthermore, the significance of these features aligns with established structural models, such as Merton (1974), wherein the assessment of default risk hinges on the relationship between a firm's equity value and its level of liabilities. Consequently, these influential features bear relevance to foundational models of credit assessment and underscore their pertinence in the present analysis.

**Figure 5:** Comparison of classification error: all features and selected features




While some features may not exhibit significance as pronounced as the above, they hold potential importance in the predictive model. For instance, the positive correlation between retained earnings and a firm's age suggests that these variables warrant closer consideration. The process of feature selection aims to identify the most influential predictors and, in this context, those exceeding a predefined threshold of 0.7 merit inclusion. Subsequently, a novel classification ensemble is trained to utilize solely the selected highly important features, and its classification error is subjected to comparison with the error derived from the preceding classifier employing all available features. This comparative analysis serves to illuminate the performance discrepancy between two distinct predictor sets, denoted as "all features" and "selected features", respectively. The aim is to discern whether a refined feature selection strategy can enhance classification accuracy and to what extent these additional features contribute to model refinement.

Figure 5 presents a comparison of classification errors between using all features and utilizing selected high-importance features. The classification accuracy exhibits minimal deterioration when less crucial features are excluded from the ensemble, indicating that featuring selected predictors is suitable for subsequent predictions. The process of feature selection can be time-intensive when the initial set comprises a multitude of variables. However, its success hinges on a judicious blend of quantitative tools and the discernment of the analyst. The variable importance measure, thus, serves as a mechanism of ranking to assess the relative impact of each feature by evaluating the extent to which random permutation of its values affects the predictive accuracy of risk classification.

The method discerns features that significantly contribute to predictive power. As indicated in Figure 5, the selected features do not perform better than no feature selection. This can be caused by information loss due to the reduction in dataset dimensionality resulting from the exclusion of specific features. The complex interactions of the features and some non-selected features that can be relevant to these interactions make the predictions less accurate. When dealing with two strongly correlated features of importance, both may receive high ranks in the analysis. In such cases, retaining just one of these features may suffice for accurate predictions, but this determination may not solely be derived from ranking results. Under this situation, one may need to consider additional examination of feature correlations or expert judgment. Consequently, while quantitative tools are useful in feature selection, the informed judgment of the human analyst remains an indispensable component of the process.

**Figure 6.** Normalized confusion matrix in percentage: decision tree

| TRUE CLASS | 1-low | 2-medium | 3-high | Accuracy | Error |
|---|---|---|---|---|---|
| 1-low | 62.9% | 28.6% | 8.6% | 62.9% | 37.1% |
| 2-medium | 10.2% | 59.4% | 30.5% | 59.4% | 40.6% |
| 3-high | 2.3% | 18.5% | 79.2% | 79.2% | 20.8% |
| | Predicted classes | | | | |

Once the model parameters have been determined, the classifier can be saved for future sessions when it is necessary to rate new clients. To predict the credit rating for new data, the "predict" method is invoked on the classifier and yields the predicted class and the associated classification score as key outputs. Among them, the classification scores provide insights into the degree of confidence associated with the predicted ratings.

Figure 6 presents a normalized confusion matrix utilizing the decision tree method. The accuracy rates for low and medium classifications stand at 63% and 59%, respectively, which are marginally lower than those achieved through the boosting method. Conversely, the accuracy rate for high credit ratings reaches 79%, surpassing the corresponding rate achieved using the boosting method. These insights underscore the nuances in classifier performance across different credit rating categories.

## 5. CONCLUSION

This study evaluates the effectiveness of decision forest and boosting algorithms in predicting credit ratings. By leveraging financial ratios as input variables, various machine learning tools in MATLAB are used in this study. Using the quarterly financial data of 2,029 credit ratings in 12 industries that were rated by five different agents between 2010 and 2016, we first reclassify the ratings to avoid the issue related to too few observations.

The empirical findings demonstrate that both these methods exhibit considerable performance but may be superior to each other in different areas. Boosting algorithms method exhibits accuracy rates of approximately 67% across the credit rating categories. The random forests model generates lower accuracy rates for low and medium classifications than the boosting method, but the accuracy rate for high credit ratings reaches 79%, more accurate than the outcome using the boosting method.

This study exemplifies how to apply appropriate machine learning models in forecasting default risk by using financial data. Furthermore, we show the usefulness of both methods exhibiting robustness when handling noisy data as they expedite training with large datasets and enhance the interpretability of the findings. The results can be useful to practitioners aiming to integrate machine learning algorithms into credit rating prediction tasks.

**REFERENCES**

Altman, E. 1968, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy" Journal of Finance 23:4, 589-609

BCBC, 2005, "Studies on the validation of internal rating systems," Basel Committee on Banking Supervision, Bank for International Settlements (BIS), https://tinyurl.com/5dbcxtrk

BCBS, 2006, "International convergence of capital measurement and capital standards: a revised framework," Basel Committee on Banking Supervision, Bank for International Settlements (BIS), https://tinyurl.com/4tnbm7wh

Loeffler, G., and P. N. Posch, 2007, Credit risk modeling using Excel and VBA, Wiley Finance

Hastie, T., R. Tibshirani, and J. Friedman, 2008, The elements of statistical learning, Springer

Merton, R., 1974, "On the pricing of corporate debt: the risk structure of interest rates," Journal of Finance 29:2, 449-470

# A SMART TOKEN MODEL FOR NATIVE DIGITAL ASSETS

**IAN HUNT**[1] | Buy-Side Industry Consultant and Adviser

## ABSTRACT

Digitalization is not about doing what we do now, but with slightly better technology. It is an opportunity to do something very different, which is much simpler, much cheaper, and much better. A number of factors are preventing us from taking, or even perceiving, that opportunity. We are held back by the popular perception that digital assets are questionable and shady, and by regulatory uncertainty over the treatment of digital assets, as well as by our own unwillingness to think beyond the current financial ecosystem: we find it hard to accommodate the degree of change that would enable us to maximize the benefits of digitalization. This paper explains the radical potential of native digital assets to create a single, simple issuance and transaction model across all financial assets. This would deliver a dramatic improvement in client outcomes and in the flexibility of investment products, along with an equally dramatic reduction in cost and risk in the way that we deliver those products. It would enable regulation to be far simpler, but more effective. It would allow us to roll back the surging tide of complexity in the infrastructure, management, and regulation of finance. This model will be adopted soon in a forward-thinking jurisdiction. All other markets, ploughing on with our current overweight, over-complex, and heavily regulated financial ecosystem, will become uncompetitive: they will have no choice but to follow.

## 1. INTRODUCTION

Forty or so years ago, Nasdaq was created by disaffected members of the New York Stock Exchange. They wanted an electronic market and had become impatient with the conservative management of the New York Stock Exchange, who clung onto the traditions of open outcry trading. Instead of persisting in their attempts to motivate change in the NYSE, the modernizers created a new venue. Nasdaq was an electronic market from day one, offered issuance and trading at a lower cost, was a more transparent venue, and was easier to integrate with participant platforms.[2] As a result, Nasdaq achieved critical mass, its volumes exploded, and it became a well-regarded and competitive trading venue. The NYSE had no choice but to react.

We are at this point again, but not just with trading. The traditional models of asset issuance, asset servicing, asset sourcing, transaction management, and trade settlement are ripe for reinvention. They now have a radical, and better alternative through native digital assets. We need another new venue, but this one will deliver a whole new way of representing assets and transactions, and a completely new market infrastructure. This paper describes the smart token model for native digital assets that delivers this transformation.

The press is full of stories about blockchain, crypto, decentralized finance, bitcoin, distributed ledgers, tokens, etc. Even the quality press often confounds these constructs, and creates an undifferentiated, shapeless mass that is loosely labeled as “digital”.

[1] In creating this paper, I have had the benefit of reviews from ten senior industry players with deep experience in investments: David Axson, Keith Bear, Rich Fox, Adrian Grimshaw, Ian Hutchinson, Mark Harrison, Ben Lucas, Keith MacDonald, Atul Manek, and Jason Webb. I am grateful to all of them all for their time, their insights, and their honest critique, as a result of which the paper has been much improved, and my excesses have been curbed to some extent.

[2] It is no coincidence that Nasdaq is now building an exchange for tokenized assets.

Regulators struggle to keep up, many taking differing views on how to categorize these new and threatening phenomena. Some jurisdictions, like Singapore, actively encourage innovation; others suppress the innovators. Some cultures pile in enthusiastically, like India into crypto, while others spectate nervously. The lack of global consistency and effective regulation gives the impression of a "Wild West" where anything goes. The digital market seems immature and chaotic.

This paper sets out to demystify the digital world, to give a straightforward account of its various components, and to show the enormous potential of digital ledgers[3] and smart tokens to deliver a much better and simpler financial ecosystem.

## 2. THE FINANCIAL WORLD AS WE KNOW IT

### 2.1 What are financial services for?

If we are to discuss the transformation of an industry, then it is a good idea to have a clear understanding of what that industry does before we start.

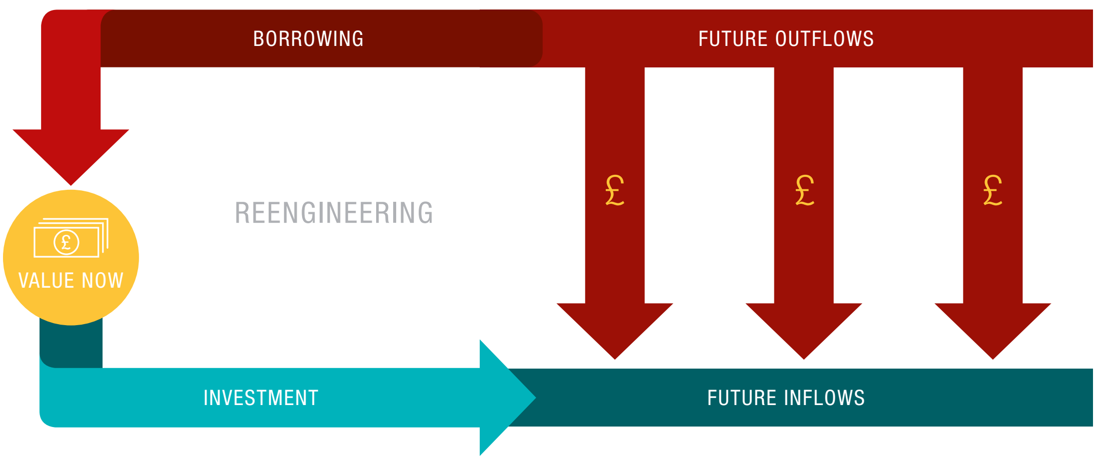
In 2021, the contribution of the financial services sector to the U.K. gross national product (GNP) was 8.3%.[4] In London, in 2020, it was 19% of economic output. There is a huge amount of activity behind this large contribution, and substantial related revenues accrue to banks, asset managers, service providers, etc. (£173.6 billion in 2021). There are also very substantial costs, driven by the extensive activity (and often wide profit margins) in the sector, which are borne by the consumers of financial services.

The range of outputs from the industry is very extensive: there are products in securities financing, collateral management, investment banking, payments, fund management, derivatives clearing, etc. The number of commercial entities is remarkable too: a small selection from our industry might include transfer agents, clearing houses, asset servicers, central counterparties, broker/dealers, asset managers, central settlement depositaries, custodians, fund accountants, market-makers, and authorized corporate directors. You could be forgiven for thinking that what we do in the finance services industry is extremely complex, requiring the most brilliant minds and the most sophisticated infrastructures. You would be wrong. Actually, the central purpose of what we do is very simple.

**Financial services simply reengineer current pots of value into and out of future flows of value. And that is all that they do.**

Investors are people or entities that have current value that they are willing to forego, in order to receive future flows at times when they need them more, and when (hopefully) they will be larger in total than the current value that they are prepared to give up. Borrowers are people who need current value, but do not have it, and are prepared to commit to deliver (probably larger) value flows in the future, at a time when they expect those value flows to be easier for them to deliver, in order to have value now.

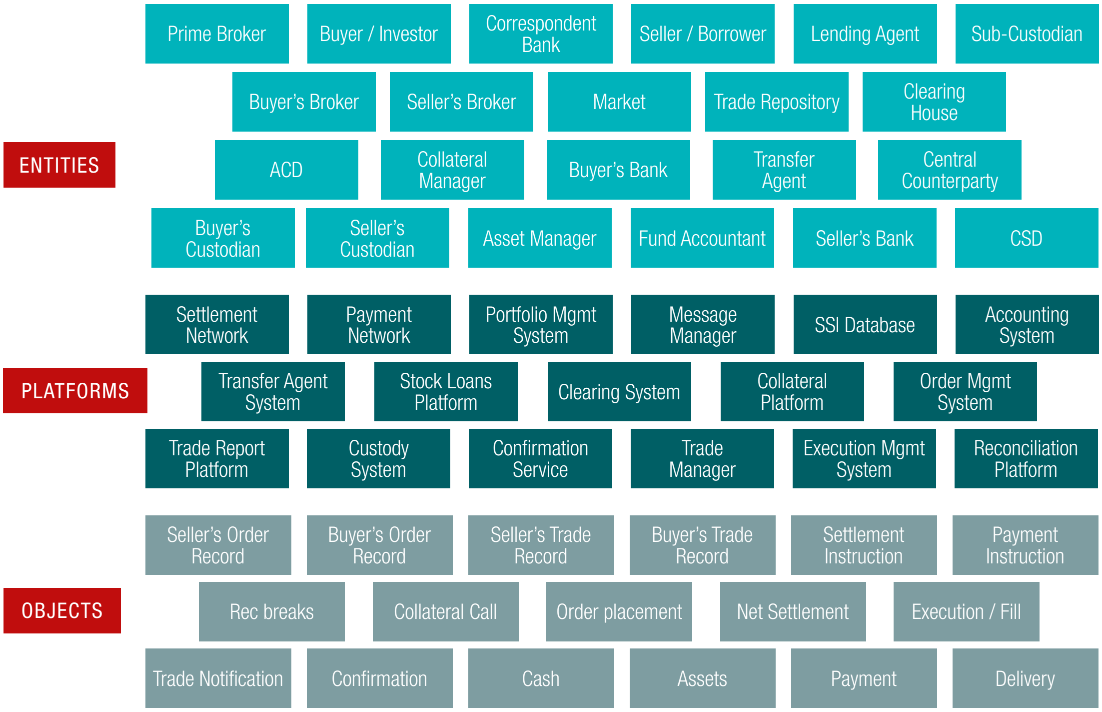
**Figure 1:** What financial services do




[3] A digital ledger is a network of nodes, representing value, ownership, and transaction in a purely digital form. It is often implemented as a decentralized network, where data is replicated at each node.

[4] Financial Services: Contribution to the UK Economy – House of Commons Library, September 2022.

**Figure 2:** The entities that we have created for one asset class




It does not matter whether the borrowers are capital issuers raising finance for their businesses, or individuals taking mortgages to enable them to buy their homes. It does not matter if the investors are institutions seeking to fund future pension payments or individuals putting money aside for a rainy day. It is all about pots of value now, future flows of value, and the ability to turn one into the other.

The entities that actually matter are very explicit in this model too: there is the borrower, who commits to deliver future flows against the receipt of a current flow, and the investor, who agrees to receive future flows in consideration for the delivery of a current flow.

**Every other entity only deserves its place in the ecosystem if it materially assists these primary participants to achieve their objectives more effectively and / or more efficiently.**

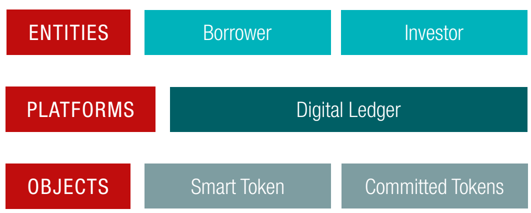
## 2.2 What conventional operating models look like

In the conventional world of financial services, every asset class has its own operating model, each of which involves a set of (often regulated) entities, multiple platforms, complex processes, heavy interactions, and burdensome regulations.

For example, in an equity operating model for a fund, depending on the nature of the investor and the nature of the trade, there may be well over 20 entities involved, each running their own platforms, and manipulating scores of data objects throughout the transaction lifecycle. Even a simplified picture is daunting: many capable industry professionals only ever comprehend a small part of it. If we added the processes that each entity carries out, and mapped the interactions between them, then the visualization would become too complex to be meaningful. And the really worrying fact is that there is one of these labyrinthine models for each conventional asset class. It is a mess.

By contrast, with native digital assets and smart tokens, we can build a single operating model that is strikingly simple, and requires far fewer entities, platforms, and objects. Even

**Figure 3:** The entities that we need in a digital ecosystem



better is that the simple operating model works for all asset classes, current and future, not just for one. In this paper, we will see how that is possible and how it is practical to achieve.

## 3. TOKENIZED ASSETS AND CASH

### 3.1 Tokenizing conventional assets

There is a lot of noise about tokenization. Multiple projects claim to create digital assets by tokenizing bonds, commercial paper, loans, and equities, as well as precious stones and real estate. They don't.

The truth is that there is nothing digital about a tokenized asset, except for the digital token itself that is the marker of ownership. Title to the asset is digital, and exists as a token on a digital ledger, but the underlying asset remains resolutely conventional. That includes its legal status, its terms and conditions, its cash flows, its operating model, the regulated entities required to manage it, the processes and interactions that those entities engage in, and the regulations that dictate what they do and how they do it.

R3,[5] in their own blog on tokenized securitization,[6] say that "Tokenization refers to converting an asset to a digital form represented by a token." Greengage similarly define tokenization as the "transformation of the rights to an asset into digital form (tokens)."[7] The obvious fact is that the words "conversion" and "transformation" imply that there is something to convert and transform in the first place: that something is a conventional asset, with all of its implications and baggage.

UBS recently issued a bond on the SIX Digital Exchange (SDX),[8] and issued the same bond in parallel on the SIX SIS, the conventional central settlement depositary (CSD) of the Swiss market. The dual listing was heralded as establishing "a migration path for the market to move from issuing traditional securities to issuing natively digital securities," and the SDX-issued security was described as a "native digital bond". It is nothing of the kind: it is just a bond whose existence is certified, and whose ownership is evidenced by a digital token, rather than by entries in the records of a CSD and a custodian. The conventional bond is collateral for the token, which has no inherent value without it.

This is not to say that tokenization is worthless, or that it is any kind of illusion. It is not, and tokenizing a conventional asset, whether or not it is a financial asset, has some demonstrable benefits. Tokens are generally easier to move around on a ledger than assets and cash in the conventional world. Tokens are easily fractionalized and traded, so ownership can be shared widely without too much pain. Having cash and assets tokenized on the same ledger enables "atomic settlement", where the payment (in cash tokens) and the benefit (in asset tokens) are locked together in the exchange: it is perfect delivery versus payment (DVP), and reduces settlement risk for both parties to a trade.

### 3.2 Tokenizing conventional cash

Tokenizing cash is similar to tokenizing assets: if the token is a marker of ownership of some conventional cash (or cash-like assets, such as Treasuries or Gilts) that exist outside of the ledger, then the token is just a claim, or an entitlement, to that pool of conventional cash. It is not anything like native digital cash.

Whether the token represents title to cash or title to assets, there is some conventional collateral, outside the ledger, that determines the value (and risk) of the token. This has implications for the parties and roles that have to be in place to ensure that the tokens have the value that we think they have. We need:

1. A safe-keeper of the off-ledger assets/cash. Obviously, collateralized tokens will lose value if the underlying collateral is lost, stolen, or turns out to be fictitious.

2. An issuer/redeemer of the tokens, charged with maintaining the relationship between the tokens and the collateral. The tokens will lose value if they are issued in excess, and gain value if issuance is short: the linkage between the value of the token and the value of the underlying is lost in either case. Someone needs to make sure that does not happen.

These parties are the equivalent of custodian and banking entities for assets and cash respectively. Hence, conventional intermediaries are required where tokens are backed by conventional collateral. Indeed, a recent paper from McKinsey & Co. highlighted just how many intermediaries can be required to operate collateralized tokens: they mention eight. [9]

[5] R3 are the developers of the Corda distributed ledger platform targeted on financial applications.
[6] https://tinyurl.com/38j4ec7x
[7] https://tinyurl.com/tjtmnj9y
[8] https://tinyurl.com/yck9yaj2
[9] https://tinyurl.com/hrc47shh

The truth is that tokenized assets are half-in, half-out of the digital world: they have digital ownership, but conventional underlying collateral. Their value and risk are derived from the off-ledger collateral, not from the digital part: the tokens themselves have no inherent value. Tokenized cash is the same: it takes the form of collateralized stablecoins (or tokenized deposits), which are tokens on a digital ledger, whose value is (or should be) pegged to an external currency or cash-like asset.

While they themselves are digital, the value of the tokens is generated by a pool of off-ledger, very conventional cash. The digital cash tokens are just evidence of entitlement to a scoop out of that conventional cash pool. The values of the tokens should be stabilized by the collateral, and track the collateral valuation precisely. However, this depends very much on the quality of the intermediaries, the rigor with which they perform their roles, and the market's confidence in them.

According to Deloitte, collateralized stablecoins will be accepted as a means of exchange by 75% of merchants by 2024.[10] The same is expected for their close relatives, tokenized deposits. According to the G7 finance ministers and the U.K. Government, collateralized stablecoins are going to be brought within the regulatory perimeter, and so are seen as officially respectable. PayPal clearly expect this to be the case, and have launched one: PayPal USD.

A recommendation from the All-Party Parliamentary Group Inquiry into Digital Assets and Crypto states that "the Regulators will need to insist that stablecoins are backed by high quality assets, and ideally by fiat currency, to provide trust and confidence in any new form of payment."[11] This insistence is a reaction to the collapse in value[12] of some prominent "algorithmic" stablecoins (like TerraUSD) that, rather than being collateralized by a pool of conventional cash, have their value controlled by a throttle on supply and demand. As the price rises above the reference currency, more tokens are issued to bring the price down, and vice versa as the price falls. These algorithmic stablecoins are not seen as reputable at all. This is understandable, but ironic, as the value of our familiar national currencies is controlled in more or less the same way.[13] The difference is that they are controlled by central banks, whom we supposedly trust, rather than by algorithms run by crypto firms, which we do not.

## 4. NATIVE DIGITAL ASSETS AND CASH

### 4.1 What "natively digital" means

Tokens do not have to be collateralized to have value: truly native digital assets are assets in token form on-ledger that are uncollateralized. Native digital cash similarly is cash that is not collateralized by any off-ledger conventional cash (or near-cash) assets. They both exist wholly and exclusively on the digital ledger, and their values and risks are not derived from anything outside the ledger.

Native digital cash is like conventional cash, in that it acts as a denomination of assets, as a means of exchange, and as a store of value. We are all familiar with cryptocurrencies, some of which are widely available and can act in these ways, but currently their volatility limits their usefulness. While they are widely held as a store of value (420 million people hold crypto globally)[14], their most common active use is for speculation.

Algorithmic stablecoins are another example of native digital cash, and were designed to address the problem of volatility. However, well-publicized failures have limited their popularity too. The white knight for native digital cash (subject to concerns on government surveillance and unlimited supply) is "central bank digital currency" (CBDC), which is uncollateralized, but as stable as the equivalent conventional currency.

Later, we will see what potential these assets and cash have in practice, how they behave, and what they enable us to do. They are much more interesting than their tokenized, collateralized relatives.

### 4.2 Why we like things to buy, sell, and hold

Laws and regulations wrap up our familiar asset types in concrete frameworks, and dictate how they are issued, owned, safekept, and traded. Each asset class has its own unique operating model (how we hold and trade them), and its own bespoke issuance model (how we create them). As a result, we have very fixed ideas about asset classes, and we see them as unique, distinct, and persistent.

The philosophy behind this is bizarre, but comfortable: we like things, and we like them to be sharp-edged. We are happy to think that, when we transact, we are buying a coherent thing. When we make a purchase, we want to have something tangible to keep, to put on our shelf, to hold in our portfolio.

[10] https://tinyurl.com/2me8fbea
[11] https://tinyurl.com/2hp6vj3n
[12] The industry term for a stablecoin losing its link to its reference currency is "unpegging".
[13] Clearly, central banks have other levers over the value of their currencies too, including interest rates.
[14] https://tinyurl.com/mv2cw3z6

When we sell, we are comfortable that we take the thing that we are selling off our shelf, and give it to the purchaser to put on theirs. We use the language of things, and talk about "delivery" from the seller to the buyer. The idea that we might be transacting without buying or selling anything coherent or tangible is very awkward. So we go along with the more comfortable allegory, and even in the context of digital assets, we cling onto the language associated with tangible, physical transaction.

There is a problem with this position: bonds, for example, are not coherent things at all. They are really no more than a fistful of promises – pledges to deliver flows of value at some points in future. The issuer commits to pay coupon at specified times and in specified amounts, and to make a redemption payment on maturity. All or any of these pledges may be fulfilled, or they may not be.

We treat bonds as if they are a single thing to own, have a single atomic value, represent a single exposure, and suffer from a single level of risk. They don't, even though it is what our systems, our laws, and our regulations tell us. Each pledge has a different probability of fulfillment, and, therefore, a different risk and a different value. The overall risk and value of a bond is a munge of these diverse pledges and their disparate probabilities. And that is before we even consider discount rates and the risk-free rate of return!

It is not just bonds that are like this. Loans are just clusters of commitments to make payments at a future date too. Swaps are commitments from two parties to make flows of value to the other party under pre-defined conditions. Options are contingent promises to deliver a flow of value to the option holder, or to accept one from the option holder. The flows of value could be, and will often be, in cash, but they may not be. For example, there are physically delivered options contracts in commodities, and equity options that deliver a flow of equities at a defined price, rather than a value delta in cash.

Any purely financial asset can be defined in terms of the flows that it commits to deliver, and is, therefore, definable as a set of contracts.[15] Each asset is just a collection of flows that the issuer has chosen to commit to, and that the recipient has chosen to receive. Hence, we can construct any conventional financial asset from flow commitments. However, looking at that the other way around, and starting from flows, rather than from conventional assets, we can clearly define whatever assets we like: the assets that we can construct are not limited to those that belong to any currently recognized asset class. We can build whatever assets are useful, and the asset class straitjacket dissolves away.

If we start from conventional assets and tokenize them, then we are stuck with our conventional assets. If we start from flow commitments, then we have the opportunity to define any asset that is useful, unconstrained by current asset classes. This presents a huge opportunity for innovation, increased asset flexibility and liquidity.

### 4.3 A single issuance and operating model across assets

In conventional finance, there is a separate issuance and transaction model for each asset type, along with a slew of regulated entities required to engage in issuance and to effect transactions for that asset. Law and regulation mirror and solidify the party lines between asset classes, and the volume of regulation is multiplied by the number of distinct asset classes. Systems tend to be asset class-specific too, and even where there are cross-asset platforms, their high complexity is driven by the need to accommodate the diverse practices and regulations of the different asset silos. This is both highly inefficient and costly.

We have established that the fundamental building blocks of financial assets are commitments to future flows (essentially pledges), that all purely financial assets are just clusters of pledges, and that the pledges allow us to dissolve the boundaries between asset classes.

**This presents us with an outstanding opportunity: we can have a single issuance and operating model across all financial assets.**

This fact has the most powerful possible consequences. Systems can be less complex and less disparate, as they only have to handle one operating model. Regulations can be simpler and less extensive, as what they are governing is simpler and less diverse. New flow commitments, and, therefore, new asset types can be created more quickly, more cheaply, and with lower risk: by definition, we will not need to develop new system capabilities to handle new flows, as they will be issued in accordance with the same issuance model, and processed in accordance with the same operating model, as all other flows. As asset classes are just clusters of standard flow commitments, the new classes can be defined and implemented without the need for system change.

[15] https://tinyurl.com/2p9mnyxe

Initiatives to date in digital finance have focused predominantly on the creation of isolated ledgers, supporting individual asset classes or products. This has created disparity and complexity, which, in turn, has led to an increasing clamor for inter-operation between ledgers. Recently, there have been some very constructive discussions on the potential for a "unified" or "universal ledger".[16] In particular, this has been seen as enabling a digital infrastructure for global payments: a "regulated liability network".[17]

The combination of a single operating model across all financial assets, with a universal ledger enabling global trading, is the most powerful possible objective: if it is practical and achievable, then it is an objective that we should pursue with all possible energy. This paper sets out to show that it is an eminently attainable goal, if we exploit to the full the potential of digitalization.

### 4.4 What is in and what is out

The potential attainment of a single operating model depends on the single issuance model for all financial assets: as clusters of future flow commitments. There is a lot of good news and there is some bad news here. Fixed income instruments and derivatives, equity derivatives, and structured products are all representable as clusters of flow commitments. Real estate, commodities, fine art, and jewelry are clearly not. Real estate may generate income, which can be represented as a set of flow commitments. However, the asset itself remains resolutely physical (unless it is in the Metaverse!). It can be tokenized, but the collateral stays physical, and exists distinctly outside of the digital ledger.

The holder of a bond (or a loan, or an option, or commercial paper, or a structured product, etc.) has a continuing relationship with the issuer that determines the risk of the asset, and, therefore, strongly drives its value. That relationship is one of pledge. The issuer makes the commitment to deliver flows of value to the holder; the value and risk are driven by the solidity of the issuer, and by the triggers, conditions, and timing of the flow.

Surprisingly, an equity is more like a piece of real estate, or a Picasso, than it is like a bond. An equity holding delivers a share of ownership in, and rights over, a company, and the company is not a digital entity. It is not exactly physical either, although it may have physical assets, but it is definitively not digital. The only ongoing commitment from the issuer of an equity to its owner is to pay dividends, at a loosely defined frequency, and in an amount specified by the dividend declarations of the company (which may be zero).

Beyond a flow of dividends (shareholder perks and voting rights notwithstanding), there is no continuing pledge relationship between an equity holder and its issuer. The main investment reason for holding an equity is the expectation of value growth. In reality, this is a hope, rather than an expectation. It is a hope of a future flow with an unknown probability, of an unknown magnitude, from an unknown party, at an unknown point in time. With an equity, no one promises you anything in respect of value realization: it is not a purely financial asset as a result.

Despite their very different and entity-heavy operating models, funds are often categorized as equities, but this is misleading. Their attributes as financial instruments result in conventional, principal-traded mutual funds being much more like bonds than like equities. Their values float with the value of the underlying assets (diluted by costs, profit margin, and liabilities), and (depending on the fund structure) the investor may own a slice of the assets. However, with a principal-traded fund, there is always a continuing relationship between the issuer and the holder. The issuer pledges redemption at whatever the fund price is at the time that the holder redeems, and it is the issuer that delivers that value flow. Sometimes, an income flow is pledged as well, and again it is the issuer that delivers the flow. Hence, a conventional, principal-traded fund can be viewed as a purely financial asset.

Additionally, and usefully, some other elements of financial services that we do not think of as assets can also be managed within the single issuance and operating model.

An order is just two pledged flows, back-to-back between two parties. Execution is just the delivery of those flows. An indication of interest in the market is just a pledge that an issuer would like to make, but has not made yet, and would like potential recipients to know about. Outside orders and trading, a corporate action is just a set of committed flows, contingent or otherwise, from the issuer of an asset. Income is a straightforward commitment made by the issuer to the recipient. No one needs to calculate entitlements – the commitments themselves evidence entitlement. Asset servicing, liquidity discovery, and order management are delivered by the same operating model as trading and settlement.

[16] https://tinyurl.com/mr4d335y. There is no implication that the universal or unified ledger is a single physical network.
[17] https://tinyurl.com/4b9n3xp2

### 4.5 A reprise on things

It is clear that there are some things that are physical, and that we can buy and put on our shelf. We have already discussed jewelry and fine art as examples that are definitely not digital. There are other things that we can buy, but that are immobile, like houses and office blocks: they are definitely physical things, but they do not come to us – we go to them. Then there are off-ledger things that are not physical, but are definitely not digital either, like companies, clubs, and governments. All of these things can be tokenized, to bring their ownership, or title, onto a digital ledger, but they cannot themselves be digitized: they are not purely financial entities.

Purely financial assets (and other financial elements), which comprise nothing but clusters of promises, are not coherent things at all, and we only confuse matters by trying to treat them or represent them as atomic entities. The fact that it is convenient from the perspective of property law and regulation is not a good enough reason to deform the representation of financial assets. We should represent them exactly as they are: as sets of commitments to future value flows. In this form, they can be represented on-ledger as wholly digital entities – in token form. These are native digital assets.

An immediately obvious advantage of a native digital asset is that it does not require the same intermediation as a tokenized conventional asset. It does not need a safe-keeper for the off-ledger reference asset, because there isn't one; and it does not need an intermediary to control issuance and redemption of tokens, because every participant can issue and burn their own.

**"Things", and tokenization, are a last refuge, where a fully digital entity cannot be created. Wherever an asset can be created in a natively digital form, we should create it in that form – it is much, much more useful.**

## 5. THE SMART TOKEN MODEL FOR NATIVE DIGITAL ASSETS

### 5.1 A fully digital ecosystem

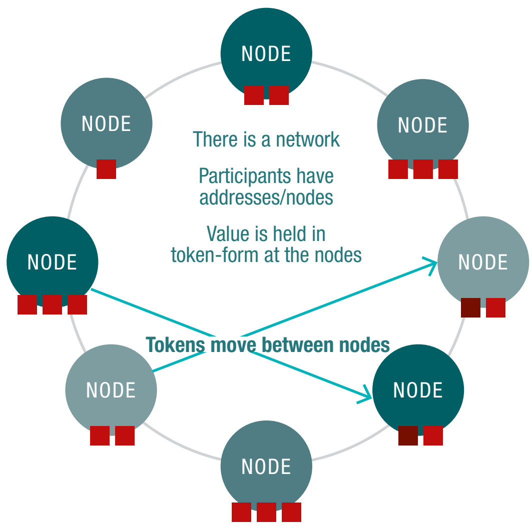
From the reasoning above, we can see what an appropriate representation of assets and transactions looks like: it is a model that closely reflects the true nature of assets, and of the flows that they commit. The transactions are the flows committed, or they are flows in the asset itself. But how would we implement that model in a purely digital ecosystem for financial services?

In a wholly digital financial ecosystem, all representations of value are in token form on a digital ledger. Tokenized assets and tokenized cash do exist on the ledger, but they are not the headline acts. Alongside the tokenized, collateralized forms of value, there are native digital assets and native digital cash, which are also tokens, but are wholly self-contained: they do not depend on anything external to the digital ledger to give them value or to determine their risk.

We have already seen that, wherever a native digital asset can be created, we should create it, rather than just tokenizing title to an off-ledger conventional asset. However, where an asset is physical or necessarily non-digital, like a building or an organization, then we should tokenize it in a collateralized form, to at least bring its ownership within the digital ecosystem. Hence, our purely digital ecosystem contains both collateralized tokens and uncollateralized, native tokens.

Every pot of value in a digital financial ecosystem takes the form of a set of tokens at an address (or "node", or "wallet") on the ledger. Every transaction is just a flow of tokens from one address to another address on the ledger.

**Figure 4:** Not much happening in a purely digital ecosystem




Early on in this paper, we established that the only real purpose of financial services is to reengineer pots of current value into and out of future flows of value. We have now seen that it is appropriate to represent purely financial assets as commitments to future flows of value, and to represent pots of value as addresses on the ledger where current value is held.

Both can be wholly digitized. In doing this, we have created a model that is not just true to the nature of financial assets, but also directly reflects the two objective deliverables of financial services: current value and future flows. The fit is remarkably close.

It does not matter what the tokens are that are held at a ledger address, or that flow between addresses: they can be tokenized conventional assets or cash or they can be native digital assets or native digital cash. Many trades will comprise a flow of asset tokens one way and cash tokens the other, in a conventional "cash for asset" transaction – but they do not have to be. Asset against asset trades, cash against cash trades, and free deliveries of assets or cash are all easy to represent in the fully digital world.

In this context, the nature of native digital cash is clear: it is a means of exchange and a store of value that is recognized and accepted within the ecosystem as such, and which has whatever value the participants in the ecosystem give it. We are familiar with true cryptocurrencies in this context.[18] There is no pool of conventional cash behind Litecoin or Bitcoin. More interestingly, central bank digital currency (CBDC) is also native digital cash, and is clearly coming down the track at us. There is no pool of collateral behind it.

The existence of CBDC, alongside regulated, collateralized stablecoins/tokenized deposits, will increase the frequency of digital trading (i.e., trades that are digital asset versus digital cash) by orders of magnitude. CBDC will be more acceptable as a means of exchange, as a store of value, and as a unit of account, because we will trust it for the same reason that we trust conventional cash.[19] CBDC is just a form of cash, and is managed and stabilized by a central bank: we have no more or less reason to trust it less than any other form of fiat currency. Stablecoins and tokenized deposits, similarly, if they are genuinely 100% collateralized, their underlying assets are safe-kept, and their issuance is controlled 1-to-1 with the collateral pool, are deserving of our trust as a means of exchange.

True cryptocurrencies are native digital cash too; however, being backed neither by collateral nor by a central bank, they are less likely to be trusted for widespread transactions. As we have already seen, their use is primarily, but not exclusively, for speculative investment, and seems likely to remain so. Other cryptos, known as "utility coins", have value only in a specific blockchain, and are useful exclusively in that context. They are forms of digital cash too, but not really relevant here. As they are "currencies" only within their own local ecosystems, they are like chips in a casino.

In a purely digital ecosystem, native digital assets can only be one of two things. The first is a non-cash asset that is purely digital and has whatever value the participants in the ecosystem give it. Good examples are "non-fungible tokens" (NFTs) that are purely digital artefacts and do not represent title to an off-ledger asset. Damian Hirst's notable collection of art "The Currency" originally comprised physical and digital artwork (i.e., NFTs) in parallel. After a year, the owners had to make a decision between the physical work and the NFT, and destroy the other. As a result, there are 4,851 Damien Hirst NFTs out there that are uncollateralized (or at least, not by physical artworks).

There is legitimate interest in NFTs and no particular reason why digital-only artefacts should not have value. However, uncollateralized NFTs are not going to transform the financial services world. The other form of native digital assets, which are tokens representing commitments to future flows, can, and will, change the world. These allow us to represent any purely financial asset as a collection of tokens. That is not just the financial assets that we are used to, and exist within current asset classes, but any financial asset the we find useful to hold or to trade now or in future. And this type of native digital asset also unlocks the door to a single issuance and operating model across all assets (and income, and corporate actions, and orders, and executions, and indications of interest, etc.).

**Big step number one is to embrace what native digital assets really are – they represent entitlements to future flows of tokens, not rights to conventional assets.**

## 5.2 What kinds of tokens do we need?

Summarizing the points above, it is clear that we need four kinds of tokens in our purely digital world. These reflect two fundamental dimensions: the dichotomy between tokens that are collateralized by off-ledger assets and those that are not; and the dichotomy between tokens that represent cash and those that represent assets. This quadrant, with examples, is presented in Figure 5.

[18] "True" implies that they are currencies that can be used for transaction generally, rather than just within a local blockchain environment. Hence, their ecosystem is widespread.

[19] There are concerns over the fact that CDBC is programmable, which could facilitate government control and surveillance. These concerns are genuine, but are beyond the scope of this paper.

**Figure 5:** The tokens that live in a purely digital ecosystem

| | COLLATERALIZED | UNCOLLATERALIZED |
|---|---|---|
| CASH | **Tokenized cash**<br>Stablecoins – cash-backed<br>Stablecoins – near-cash-backed | **Native digital cash**<br>True cryptocurrencies<br>Central bank digital currencies |
| ASSETS | **Tokenized assets**<br>NFTs – title to off-ledger assets<br>Tokenized conventional assets | **Native digital assets**<br>NFTs – native on-ledger<br>Commitments to future flows |

## 5.3 Making native digital assets smart

It is embedded in our mindset about technology that intelligence and capability lie in business systems. The functions, products, processes, and operations of each business are coded into their business systems, in all their complexity and variety. The business system does what it does, enforces workflow, makes computations happen, and pushes dumb data and messages around like a croupier raking chips on a casino table.

This idea is so embedded that we take it wholly for granted, and it is seldom, if ever questioned. Even modern platform developments, including distributed ledger platforms, follow the same pattern: they support "smart contracts" that self-execute on the ledger, but that are wrapped up in apps that sit at defined locations, and push dumb tokens around on the ledger. The tokens are generally assumed to represent a quantity of some conventional asset, and by implication are collateralized. The tokens themselves may be programmable, and carry their own rules, but being based on conventional assets, they can only deliver tweaks on existing operating models, not anything radically new.

The world changes significantly if we take the intelligence and power away from the immobile business systems, and put them instead on the tokens: we make them intelligent and potent. This may make it sound as if the smart token is a very complex entity, encapsulating the entire range of functions of a business system. But it is not. By definition, on the digital ledger there are only two things taking place: the holding of tokens at addresses, representing stores of value, and the movement of tokens between addresses, representing transfers of value. Nothing else is happening.

It follows that the only thing that a smart, potent token can possibly do is to move tokens between addresses on the ledger: this may be moving itself, or it may be moving other tokens in fulfillment of the commitment that the smart token represents. Flow commitments by smart tokens are just active contractual liabilities: they are IOUs with the power to make the committed transfers actually happen. We will call the other tokens that the pledge on a smart token commits to transfer "committed tokens". Committed tokens can be any of the four kinds of tokens that populate the ledger.

The smart token is a token, and, therefore, lives at an address on the ledger: that address is the address of the recipient. Hence, the beneficiary of the future flow holds the token that pledges the flow, just like holding an IOU. The difference between a smart token and an IOU is that the smart token is capable of making the committed flow of value happen, without intervention from the issuer or the recipient. Once the committed flow has been made to happen by the smart token, then the smart token itself, just like an IOU, can be returned to its issuer; it can then be torn up or used again at the discretion of the issuer. The difference between a smart token and a conventional IOU is that the smart token can send itself back.

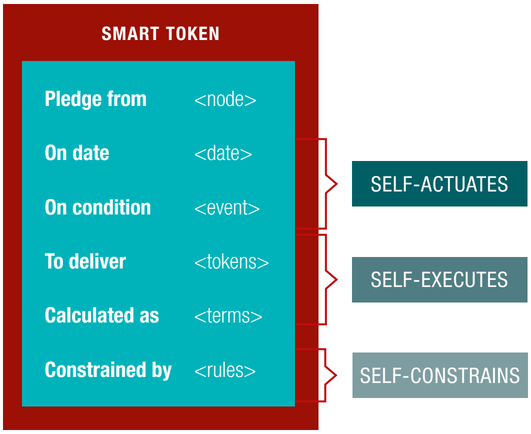
The smart token, as a token like any other, can be traded by the recipient, and can be fractionalized; so, the recipient has complete discretion over how much of it is traded on. The identity of the issuer is unchanged by the trade, and the issuer's aggregate commitment is unchanged by either fractionalization or by on-trading of the token.

If an issuer wants to issue, or a recipient wants to issue or trade a recognizable asset, like a bond or a loan, then they will pull together the tokens that represent that asset, and issue or trade them as a cluster. However, they do not have to: each token within a cluster can be traded individually,[20] as fractions, as part of the original cluster, or clustered with other tokens outside the original cluster. Hence, the resemblance to a conventional asset does not necessarily persist after its initial issuance, and the market is an order of magnitude more flexible as a result.

Each smart token only needs to have a basic set of conditions encoded to enable it to self-actuate and self-execute. It needs to know who the issuer is, what kind of tokens they have committed to deliver, how many of the tokens they have committed, when it needs to happen, and what constraints

[20] Like trading an individual future coupon with a conventional bond.

**Figure 6:** Smart tokens are surprisingly simple things




(if any) apply to the transfer. It can then do its job, kick itself into life, work out the terms of the commitment, and move the tokens. It does not need to know where to move the tokens to because, like an IOU, the token will always be held at the recipient's address.

When we make native digital asset tokens smart and potent, exciting things happen. Entitlements are not just represented in a self-maintaining form; they are implemented 100% automatically. Transactions happen when intended, in the agreed asset, in the quantity intended, and between the contracting parties, but without intervention or management from either party. The parties have already signaled their agreement to this: the issuer by issuing the token, and the recipient by accepting it. Settlement management, order management, and execution management dissolve away as a result. Asset servicing is shredded.

Every smart token is both an asset (to the recipient) and a liability (to the issuer). It can be thought of as a self-executing contractual liability, which is held as an asset by its beneficiary. As a result, asset and liability management can be much more precise, and both sides are naturally represented on the same ledger. Liability matching ceases to be an approximation of cash flows, implemented roughly in bonds and tuned with complex derivatives. Pension, insurance, and endowment managers do not need an asset platform, a liability platform, and an LDI platform to bridge the two.

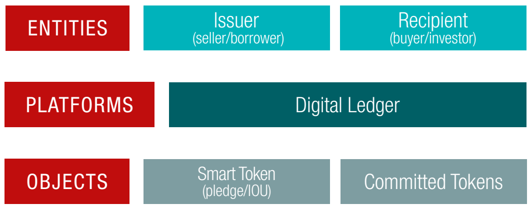
**Big step number two is to realize that native digital asset tokens should be smart, potent, and capable of implementing the commitments that they represent.**

### 5.4 What the single digital operating model looks like

When we create and work with tokens that are smart and potent, we can implement not just the single operating model that native digital assets facilitate, but we can also implement a model with self-execution at its core.

The only entities that we need in the smart token operating model for native digital assets are the issuer and the recipient of the tokens. The platform is provided by the digital ledger. The only objects required are the smart tokens themselves, and the tokens that they pledge to deliver. The difference from a conventional model for one asset class is stark and extreme.

**Figure 7:** Entities in the digital operating model




The processes required in the operating model are very limited too: six, if the smart token is issued as part of the transaction, and five if the token has already been issued and is just being traded.

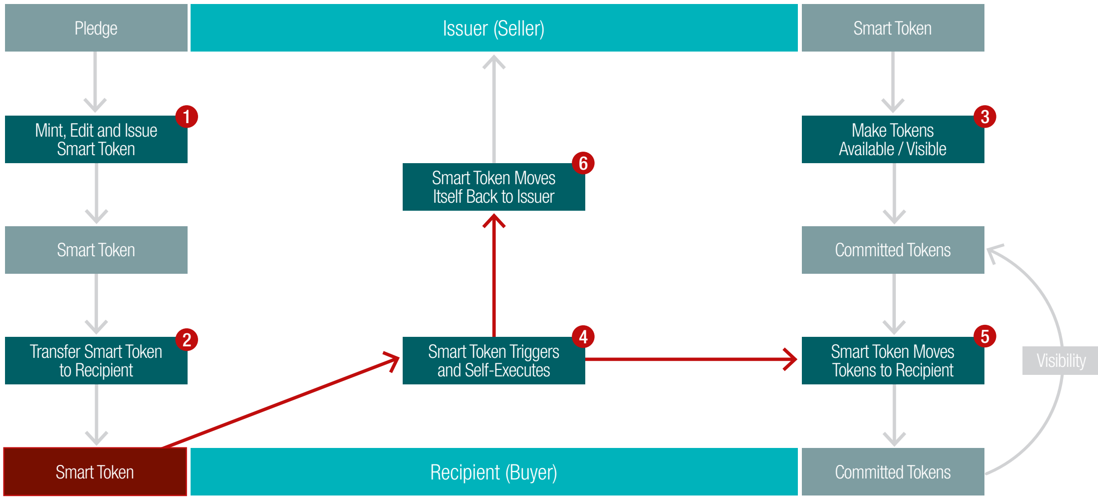
The steps are:

1. The issuance of the smart token (if the smart token is newly minted, rather than being sold on by a previous recipient).
2. The transfer of the smart token to its new recipient.
3. The earmarking of the committed tokens by the issuer, making them visible to the recipient.
4. The self-initiation of the smart token and the computation of its terms.
5. The transfer of the committed tokens from the issuer to the recipient.
6. The transfer of the smart token back to the issuer.

That is all.

Any financial asset that we issue or trade, and any derivative that we wish to represent, can be transacted by one or more instances of the same operating model. The model is largely self-executing, based on the capabilities of the smart tokens.

**Figure 8:** Not much happening in the smart token operating model




The extent of automation and simplification achieved is remarkable, and the consequences are orders of magnitude reductions in cost, operational risk, and regulation.

## 5.5 Rolling back the tide of regulation

The extent of regulation in any ecosystem is a function of the number of operating models in the ecosystem, and the number of entities, platforms, processes, and interactions required in each operating model. On top of this, we can add the sum total of the risks that these (and the infrastructure) pose to the participants, which is a function of the complexity of the system itself. In the conventional financial world, there are large numbers of each of these, and that results in a very large body of regulation. In the smart token ecosystem, there is one operating model, which is dead simple; there is only a handful of entity types, processes, and interactions, and one platform. Regulators only need to regulate what is there, and as a result, regulation shrinks from an encyclopedia to a pamphlet. It can also become much more focused and much more effective.

Some regulators and legislators argue (and flip-flop) about the classification of digital assets, desperately trying to tame the beast by locking it up in a familiar box: is crypto a currency, a security, a commodity, a digital property? A recent court ruling in the SEC's case against Coinbase and Binance has created further confusion, implying that a cryptocurrency is a security if it is sold with a contract to institutional buyers, but it is a commodity when it is traded on exchanges.[21] Others want to treat digital assets (and crypto) as a new asset class, in need of new entities, processes, controls, and regulations.[22] This is wholly wrong-headed: the opportunity presented by digitalization is to achieve a radical simplification, not an extension of complexity.

**When we implement the smart token model, we will roll back the tide of complexity in regulation.**

## 5.6 Smart tokens in funds, insurance, and pensions

Funds apparently offer choice to investors, because the choice of assets and strategies is extensive. However, in practice, their product is almost invariably the same: the value delta (net of liabilities, costs, and profit margins) between subscription and redemption by the investor. And conventional funds kick all of the risk over the fence from the fund to the investor.

If funds start to issue smart tokens instead of shares and units, then they can commit any flow that the fund can support on an asset/liability basis. Product choice and personalization become a reality, not through complex product structures, but as a direct and simple consequence of the model. If they are sufficiently capitalized, and have the appetite for risk, then funds and fund managers can use their own balance sheets to create products that their clients want. The allocation of risk and reward between the ultimate owners and users of capital can be tuned to meet their objectives and appetites, and the

[21] https://tinyurl.com/yemn9cxk
[22] https://tinyurl.com/5bjt4x3a

boundaries between the buy-side and sell-side, and between asset managers and banks can erode.

There is no need for rigid definitions of fund types – one structure suffices for all. Funds become outcome-focused, rather than just delivering whatever returns their underlying assets happen to generate. Every fund is both principal-traded and market-traded, and every fund product can be either open-ended or closed-ended within the same fund. If we apply the smart token model to funds, then every level in the fund value chain – from underlying asset management, through the fund entity, up to platforms and distributors, and into IFAs and their clients – operates under exactly the same operating model. The same relationship of outcome commitment exists at all levels.

Platforms can pass through the outcomes committed by their funds, or they can take a more active distribution role and reengineer them to create their own products to suit their clients' needs. With funds based on smart tokens, conventional collective investments, insurance, pensions, and endowments are all just different outcomes, delivered from the same structure and under the same operating model. The only difference between them are the pledges on the smart tokens held by their investors. If the pledge is a stable or inflation-adjusted flow, then it is a pension. If the flow is triggered by a loss event, then it is insurance. If the pledge is the value delta between investment and redemption, then it is a conventional collective investment.

### 5.7 What is left for business systems to do?

Growth in the scale and complexity of business systems has been inexorable, and seems inevitable. As further areas of business activity are automated, as workflows are mechanized, and as the scope of decision support and artificial intelligence spreads, so system platforms expand. The story has echoes of the universe itself, and of regulation – an ever-expanding cloud of complexity.

The smart token model transfers the work of end-to-end trade management from the business system to the tokens themselves. The tokens also take on all of the work of entitlements and asset servicing. Hence, business systems in finance lose some of their most central functions. However, the impact of a fully digital ecosystem, including smart tokens, is even much more profound than that: business systems in finance will become much smaller and much less diverse.

The root of this is the simplicity of the digital environment itself. To ensure the secure operation of the ledger as a whole, service providers will be required (among other roles) to run infrastructure, to validate digital identities, to verify that issuers can meet their committed flows, and to facilitate liquidity. However, the core entities are just the issuers and the recipients of tokens. And each participant can be both an issuer of some tokens and a recipient of others, so the core functions required to support each participant, and each address on the ledger, are the same. These are the services that center around the creation, issuance, receipt, holding, and trading of tokens. They are the services that the smart tokens cannot do for themselves: actions that require the participants to make decisions and to take responsibility.

It is a basic requirement that the owner of an address should be able to receive tokens at their address, to hold them securely, to have visibility of them at all times, and to appraise their risks and their values. They also need to aggregate these values and risks to achieve an overall view of their invested position and future inflows. However, the recipient must not be able to edit the received tokens in any way: like an IOU, the holder cannot arbitrarily change the commitment of the issuer.

The participant needs to be able to mint, edit, and issue their own smart tokens, and send them to their recipients. Those tokens may be on-traded by the recipients, and so may end up spread around the network.

Consequently, participants must be able to see all smart tokens that identify them as the issuer, wherever the tokens are, in order to verify that their aggregate commitment is what they know they have issued. This is a basic check on counterfeit or corrupted tokens.

Inevitably there is risk in coding and errors can be made: the ability to verify and audit smart tokens is critical and new systems of dispute resolution and redress will be needed. However, disputes will be much simpler to resolve: what a smart token did and why, along with liability and obligation, will be much clearer.

Tokens can be fractionalized and traded. Each participant needs the ability to trade-on the tokens that they hold, as a single token, in fractions, or in clusters, entirely at their discretion. Their business platform will support this process. The clusters do not need to reflect the shape in which they received the tokens, so they could buy a cluster representing a bond, but then sell a cluster in a wholly different shape, depending on market appetite for the tokens.

Both issuers and recipients of tokens need to identify the demand across the network for the tokens that they hold or that they may choose to issue. Issuers will invite trading by advertising flow pledges that they would like to make, exposing unissued pledges as "indications of interest"; similarly, recipients will expose the tokens that they are prepared to on-trade, so that they are visible to potential new recipients. Both parties need the capability to search the network for demand that matches their own trading objectives, to create order pledges from the matching tokens, and to issue those pledges.

Participant nodes hold only inbound tokens, but all participants can be both recipients and issuers. Many participants will, therefore, have issued tokens committing future flows, and these will be in circulation on the network. It is fundamental to the governance of the native digital asset model that the issuers of tokens should be capable of meeting the flow commitments that they have made (or have provision for alternative liquidity). There is, therefore, an asset/liability management responsibility on all issuers, and their platform will support this.

All business systems, for all participants, are of this form. Hence, in the digital ecosystem, we achieve a radical simplification, a material convergence, and a reduction in complexity for business systems. This is as beneficial as it is unusual.

**When we implement the smart token model, we will roll back the complexity of business systems.**

## 5.8 Who benefits and who loses from smart tokens?

Early initiatives in blockchain suffered from a tendency to bite the hand that fed them, threatening the business models of their own sponsors: generally, banks and custodians. The industry reaction has been to move to narrower use-cases, seeking tactical improvements within existing operating models. This has been more successful in its own terms, but will never change the world in any meaningful way.

The smart token model for native digital assets is the opposite. It is how it is, because it delivers a simple and efficient view of investment entities and market infrastructure. It is not a construct targeted at the elimination of specific existing entities in the financial ecosystem nor at disintermediation in general. It offers no tactical improvements to current operations. Inevitably, because it is so simple, there are fewer entities required to make it work, and its fundamental entities are just twofold: the issuer (i.e., the borrower) and the recipient (i.e., the investor/asset owner).

Most current roles just do not exist in the smart token operating model, but for new entrants or for existing businesses that reinvent themselves, there is much space to deliver services that enhance the operation of the ledger. This may be by operating market infrastructure, by running smart markets, by adding liquidity, by underwriting settlement, or by optimizing net flows, etc. The dramatic increase in efficiency offered by the model will lead to a broadening in participation and a growth in volume of financial markets. For those prepared to evolve their business models, there is a new and larger world of opportunity.

**The losers are easy to identify. They will be the existing players who seek to obstruct change and to protect their current revenue flows, rather than working out what value they can add in a wholly different ecosystem.**

The model is not just for business entities and institutions either. Anyone can be an issuer and/or a recipient, whether they are an individual, an institution, a club, or a business. They can all benefit from the move to outcome-focused products, from the democratization of investment, from low-cost granular trading, from improved liability matching, and from the rollback of complexity in systems and regulation. There is something in this for everyone.

## 5.9 The first step – we should create a native digital asset venue

Someone, somewhere, will be the first to do this.

Before jets swept aside piston-engined airliners, we spent fifty years incrementally improving the piston engine, adding more and more accessories and complexity to squeeze more performance out of an increasingly obsolete machine. When the jet engine was conceived, it was developed in parallel, while piston airliners continued to fly. No-one seriously tried to reengineer piston-engined airplanes into jets. Jets were much simpler, much quicker and more reliable, and new models were built from scratch to optimize the potential of the new power source. Once they became available, jets rapidly attracted the market away from piston-driven models. Piston-engined airliners are now seen for what they were – overly complex, under-performing, and now obsolete museum pieces.

The current model of financial assets and trades, and the infrastructure that supports it, has been incrementally improved over many years. It is now more performant than ever, but it is also bloated, complex, and expensive. It labors under a crushing weight of regulation, brought about by its own labyrinthine complexity and fragility. The idea that we can somehow migrate, in a controlled fashion, from this miasma into the clear air of a future state, is naïve. Vested interests will obstruct it, and the inertia of regulation and legislation will stifle it. It is a fool's errand, and we will all be dead before it happens.

What we need to do is to create something better, alongside and separate from the current model, and let it grow naturally. We can let the existing ecosystem wither over time: it will end up a museum piece too.

We need the creation of a new venue, in an ambitious and sympathetic jurisdiction. It will be a venue where native digital assets can be issued in smart token form, and where trades can be managed across assets through the common operating model. It will be quicker and cheaper to issue, quicker and cheaper to transact, and much quicker and cheaper to create new products and asset types. Transparency will grow; and while regulation will shrink, it will become more effective.

Such a venue can be created, if necessary, with a focus on a single asset class, and within a regulatory sandbox, if it has to be. But if it works for loans, it will work for swaps; if it works for options, it will work for bonds; if it works for collateral, it will work for corporate actions. Hence, the initial target is essentially irrelevant. It may grow from current token exchanges, like SDX or Archax, or it may be built by a new and innovative developer. Its creation requires just a jurisdiction that wants to facilitate change, a credible infrastructure provider, an initial issuer (so a sell-side participant), and an initial recipient (so a willing buy-side).

Liquidity cannot be created at a stroke, but like Nasdaq, the advantages of the new venue would attract issuance and trading volumes. Ultimately it would achieve critical mass, and become self-sustaining. At that point, again like Nasdaq, growth in the new model could become explosive, and other jurisdictions would have no choice but to react.

So, we have been there before. The difference is that smart tokens and native digital assets are more radical than Nasdaq, imply more fundamental change, and have deeper and more widespread benefits.

**One jurisdiction will do it; the rest will follow.**

Image generated by Adobe Firefly

## 6. CONCLUSION

The great thing about a purely digital ecosystem is that not much is going on: value is held at addresses on the digital ledger in token form. And value moves between the addresses on-ledger as flows of tokens. That is all; nothing else is happening. As a result of this radical simplicity, we can build operations and issuance models that are similarly and strikingly uncluttered.

If we start from commitments to future flows as the main form of native digital assets, then we can have a single, simple issuance model, and a single, simple operating model across all assets. We can represent familiar asset types, but we can also represent anything else that is useful to investors and to borrowers, to capital issuers and to asset owners. The hard boundaries between asset classes melt away.

When we make the tokens smart and potent, then the single operating model becomes self-executing. Settlement management, asset servicing, position management, and payments all cease to be activities that we need to resource and manage. We do not need registry or entitlement calculations either. The model accommodates assets and liabilities completely even-handedly; there is no need for separate asset and liability platforms.

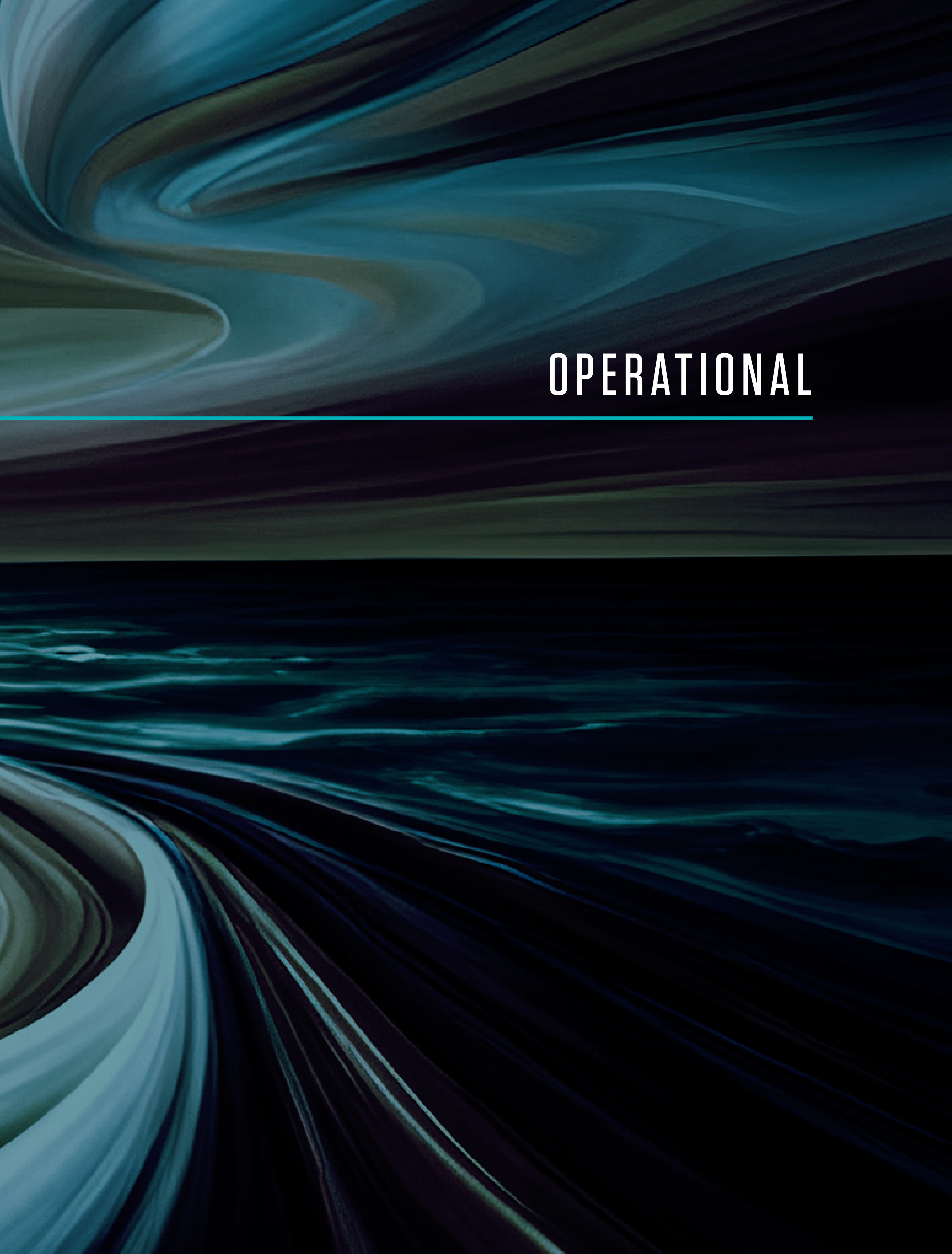
New products and new asset types can be built just by coding new smart tokens. The same operating model (and, therefore, the same tech) will support whatever is issued, so change becomes very quick, very cheap, and very low risk. In the conventional world, introducing a new asset class takes years; with smart tokens, it can be done in less than a day.

The apparently inevitable growth in the scale and complexity of business systems and regulations can be rolled back. This is not a nirvana for reactionary players railing against regulation and automation: both can be more effective and more comprehensive because their context is much simpler, and their focus can be commensurately sharper. Better, more comprehensible regulation and smaller, more manageable applications will follow.

The impact on cost and time in the industry is profound. The traditional cost models for investment funds, insurance funds, pension funds, and other financial products are transformed and aligned. The costs of issuance, transaction, technology, regulation, and compliance come down by a quantum, with multiple cost-drivers eliminated altogether. Latency and friction are eradicated, while transparency and product choice are enhanced.

Digitalization is not about doing what we do now, but with slightly better technology. It is an opportunity to do something very different, which is much simpler, much cheaper, and much better. The smart token model shows us the way to take that opportunity and to yield its benefits to the full.

# OPERATIONAL



# NETWORKED BUSINESS DESIGN IN THE CONTEXT OF INNOVATIVE TECHNOLOGIES: DIGITAL TRANSFORMATION IN FINANCIAL BUSINESS ECOSYSTEMS

**DENNIS VETTERLING** | Doctoral candidate, Institute of Information Management, University of St. Gallen

**ULRIKE BAUMÖL** | Executive Director of Executive Master of Business Administration in Business Engineering, and Senior Lecturer on Business Transformation, University of St. Gallen

## ABSTRACT

Business today is not conducted by single organizations alone but in networked designs with diverse actors. A construct where actors engage in joint value creation is called a business ecosystem. Specifically, within the context of core services originating from the financial services industry, such constructs are called financial business ecosystems. Innovative technologies and intelligent methods enable value creation in these organizational setups. To effectively participate in these ecosystems and exploit the potential of innovative technologies and intelligent methods, organizations need to develop a novel operating model. We propose a blueprint for such an operating model building on two levels of capabilities: first level capabilities that enable the exploitation of data and the number of partner relations as underlying resources of business ecosystems. The proposed second level capabilities enable the organization to engage in business ecosystems. By suggesting these capabilities, we aim to guide organizations on a targeted transformation journey and enable them to leverage innovative technology for actively engaging in financial business ecosystems.

## 1. THE EVOLUTION OF BUSINESS ECOSYSTEMS AND THE ROLE OF INNOVATIVE TECHNOLOGIES

Helvetia, a Swiss insurance company, and myky, a Swiss financial services company, are two companies that have one thing in common: both use the potential of partners to provide as complete a service offering as possible to customers. Data and its targeted use, supported by new technological developments such as AI (artificial intelligence), often form the basis for such service offerings today.

Doing business today is formed by organizations that do not work alone but try to form networks of organizations, within which forces are joined to create value for end customers by utilizing distinct capabilities and innovative technology. Prominent examples include Apple, PingAn, Amazon, and Alibaba.

One specific form of these constructs of different organizations is called business ecosystems. Ecosystems in biology describe the organisms and the interactions between them and their surroundings within an integrated system [Tansley (1935)]. In management, a business ecosystem is a unique form of joint value creation by a group of organizations [Adner (2017), Jacobides et al. (2018), Moore (1993)]. The examples of Helvetia and myky are defined as financial business ecosystems. Financial business ecosystems, more specifically, denote the structure of organizations that enable a joint value proposition to arise, whereby the core service around which the organizations group is related to the financial services industry. In the case of Helvetia's "home ecosystem", Helvetia tries to provide all possible services related to housing [Seehofer and Lechner (2023)]. To do so, they combine the offerings of a wide range of partners, from mortgage lending platforms to home-security providers [Seehofer and Lechner (2023)]. The financial business ecosystem myky supports homeowners with tools to manage their properties and their move toward more sustainable housing [myky AG (2023)]. Originally, myky was founded by an energy provider, an insurance company, and a regional bank in Switzerland. Since then, even more regional banks are backing myky, and the expansion of the service portfolio with partners in the business ecosystem is being driven forward to increase the benefits to end customers. In the case of myky, the core of the business ecosystem is constructed by a digital "house file" that serves as the point of gravity for curating the customer journey, incorporating information on the housing situation.

The evolution of business ecosystems is, to a high degree, driven by innovative technologies and the respective capabilities of the organizations within to exploit the potential of data and relations with other organizations. Among others, some examples are using AI models to create innovative customer solutions, explore new business models, or manage the value creation process among the different actors. At the current rate of development of AI applications, this trend is likely to accelerate significantly in the future. Consequently, the database as a source for the AI models becomes a critical success factor for organizational designs, such as business ecosystems and other design variants. One important thing to remember in the context of organizational designs is that exploiting the potential of technology by focusing on a single element, such as data as the basis for exploiting AI, will fall short. Rather, it is important to cover the entire operating model and identify necessary capabilities for the organization in the various areas involved.

Organizations today turn to business ecosystems because their surroundings demand a transformation of their value creation process. In the light of a holistic approach, three main topics drive this transformation: the evolvement of (1) society, (2) customers, and (3) IT innovations.

Turning to (1) societal developments, connectivity and individualization are two so-called megatrends [Zukunftsinstitut GmbH (2023)]. Almost no individual or organization exists independently, but always in connection with others. This is a development that also grounds the analysis of business ecosystems provided by Moore (1993). Technology enables

*"One important thing to remember in the context of organizational designs is that exploiting the potential of technology by focusing on a single element, such as data as the basis for exploiting AI, will fall short."*

interconnectedness – be it between individuals or organizations or both and machines. Nonetheless, individuals strive for individualization. This exemplarily manifests itself in the many different ways users of smartphones can individualize their phones (or other devices) and install apps to manage their daily lives or wellbeing. Such individualized offerings are rarely provided by one single organization – imagine Apple providing all existent applications on their own.

Further, customers (2) are increasingly demanding services or products that are complex and hence can often not be provided by one single organization [Dattee et al. (2023), Moore (2023)]. Reasons for this are, on the one hand, that customers have access to much more information on what is possible for a product or service concerning design, functionality, and pricing. On the other hand, the megatrend individualization drives the need for a unique solution tailored specifically to individual requirements. Consequently, to stay with the smartphone example, it's likely that no two smartphones have the same setup. For the financial services industry, that already leads to service offerings that are combined of service elements (also called increments) of different providers from different industries, as, e.g., in the cases of myky or Helvetia.

Evolving information technology (3) supports these developments, both from a customer perspective as well as from an organizational perspective. The modularization of services and, with that, offering single increments or individually combined increments for a joint value proposition is one important part of implementing customer centricity. In the financial services industry, this is shown, for example, in the evolution of so-called fintechs (financial technology companies), offering specific increments of a financial service solution as a single solution or in the "open banking initiative" that aims to integrate payment services into value creation processes. The occurrence of fintechs offering increments and initiatives in the direction of open banking calls for more flexible and adjustable core banking systems that possibly allow for the inclusion of the offerings of other organizations than the original bank. Standardized protocols such as application programming interfaces (APIs) efficiently support connecting different players from various industries. These allow for connecting various organizations based on prescribed conditions and foster a trend of organizations moving towards opening for collaboration. Seeing more customer value created in settings of different actors, APIs might be seen as the enabler of such joint settings in providing a way of efficient interconnection.

Today, a second important aspect connected to the development of innovative technology, such as AI models, and their application is, as mentioned before, the database and shared data underlying the value creation structure. If intelligent algorithms are to be used to generate new service offerings, propose new business models, drive hyper-automation, or orchestrate players in a business ecosystem, the "right" data must be available.

Following the holistic approach to technology-driven transformation, this article aims to develop insights into an operating model for financial services companies that consider being part of a business ecosystem and using innovative technologies, such as AI.

## 2. DIGITAL TRANSFORMATION AMPLIFIES BUSINESS MODEL TRANSFORMATION

A growing amount of data exists, underlined by the development that an increasing number of services and products are provided digitally or evolve from a solely physical to a physical and digital world – a story that certainly accelerates data growth. Exemplarily, some years ago, refrigerators had the sole purpose of sustaining goods – a rather physical service. Nowadays, the refrigerator is a connected assistant that might schedule groceries or even order them without human interaction necessary – the payment services included as an increment. Hence, data about the status of the refrigerator becomes an intervention point for businesses – something that was not possible before having the refrigerator as a data producer in the kitchen. Consequently, it transforms from a "physical" provider of sustained goods to a "digital" recommendation and processing machine for doing groceries. Additionally, devices like smartwatches produce more data every second, enabling new services by organizations, like alarming the user when detecting early signs of disease. In both examples, the potential of data is excavated by using AI to identify patterns and conduct or trigger activities such as scheduling groceries or alarming the user. On the one hand, the high amount of data created at the customer interface calls for database models that allow for storing and making this data available for joint value creation of different organizations. On the other hand, innovative technologies and respective methods, such as intelligent algorithms, are needed to exploit the potential, for example, by identifying patterns. It is probably safe to say that digital transformation and, with that, business transformation is very much driven by the ability to manage data.

Data management in business ecosystems faces several challenges. First, data is created in the individual organizations being the actors in the ecosystem, and second, data is created in the ecosystem as a whole. The respective data models need to be coordinated with respect to a logical integration, and managing data quality is key to the quality of the service solution. As the more critical discussions around machine learning applications such as ChatGPT show, it only provides "valuable" results when the data from which it sources the answers is available and correct [Dumitru et al. (2023), Yao et al. (2020)]. Furthermore, a challenge for integrated data management in ecosystems is the reliable creation of increments. If only one increment of the whole service solution fails, the entire solution is probably at stake. An explanation for this can be found in the dependency of the organizations providing increments for the overall solution [Lingens et al. (2023)]. Helvetia, for example, tries to cover a rather long and diverse customer journey from the interest in a property, financing, to its management in one single solution [Seehofer and Lechner (2023)]. They exemplarily offer the opportunity to increase the validity of the individual decision to buy a house by integrating PriceHubble, a self-owned venture of Helvetia, in their service offering. Furthermore, they offer the choice to compare and purchase a mortgage by integrating services from the organization Moneypark. For the management of a house, they include services from organizations providing burglary prevention or digital access systems. Following the vision of "Offering everything related to accommodation from one source" [Helvetia, n.d. (2023)], Helvetia here integrates the increments of various organizations into an as complete as possible service offering provided at one place. Even more organizations might be included in the overall service provision in the future. This might increase the complexity and further create the need to guide the customers through the system. Imagine Amazon not offering a search function or providing recommendations – customers would value the service's usability significantly less. To be able to guide the customers and develop the business ecosystem in general, the increments and the related data (input data, i.e., the identification of customer needs, "production" data, and output data) must be available for analysis, and use.

The complexity within the framework of business ecosystems and the fact that data plays a prominent role here, whether on the side of the "provision" of digital increments by the organizations in the business ecosystem or through the fact that more and more interaction points with customers generate more data that must be used, highlight, once more, the importance of data for today's value creation in business ecosystems.

The business transformation, driven by the digital business that ecosystems pursue, starts with developing data management capabilities and creating the right database(s). However, for organizations, it is not "just" about having data but also about organizing and using it in a targeted way. A main step in the transformation here is to break up data silos and integrate the data either in a data lake or a consistent database that serves the entire business ecosystem.

Based on the integrated data, innovative technologies and methods, such as the diverse instances of AI, can take effect. As previously mentioned, they can have an impact on the interface to the customer by providing tailored solutions, and they can have an impact on the more efficient creation of these solutions in the backend, which are provided by the network of organizations that enable their creation. To exploit this potential, at least in the business ecosystem, capabilities to use AI must be developed. The capabilities cover, for example, understanding the mechanisms behind the specific model that is to be used, being able to create a suitable business concept, or building the operational readiness in the business ecosystem to integrate AI. Operational readiness is obviously a key factor and does not only refer to the business ecosystem as a whole, but also to the individual company being part of it. Consequently, the transformation of the operating model as the main source of operational readiness is discussed in the following section.

## 3. NETWORKED BUSINESS DESIGN AND THE TRANSFORMATION OF THE OPERATING MODEL FOR DIGITAL VALUE CREATION

To understand the capabilities that must be developed in a business ecosystem to create a complex service solution based on data, it is important to understand their coordinative design.

In business ecosystems, the actors provide increments that, when connected, form the overall service offering. Due to the dependency of the actors on one another, coordination is necessary, a task that is often conducted by one organization named the orchestrator [Lingens et al. (2023)]. However, the dependency does not only create the need for coordination, but it is also a driver of why business ecosystems are a relevant modus operandi of value creation today.

To profit from the chances offered by the construction of business ecosystems, organizations have to transform their business models based on shifting their focus of interaction. In rather classical settings, differentiation from competitors by delimitation was oftentimes considered a good way for creating value. In business ecosystems, differentiation does not build on delimitation but on the provision of the best capabilities and cultivating a relationship with other actors in the construct based on both competition and cooperation (sometimes called coopetition). Due to the dependency on increments offered by other organizations for producing the overall service and maximize value creation, opening up and jointly developing the whole system is key. Data as a key resource, as well as the connection to partners, gain importance to manifest their own position within the business ecosystem. Furthermore, excavating these key resources in applying their own capabilities leads to a new form of differentiation and will create a goal-oriented position within the business ecosystem. However, what is important to keep in mind is that the focus of an organization needs to shift away from "getting the largest piece of the cake" towards "baking a larger cake together with other actors"– organizations need not focus on accumulating most of the value created themselves, but rather increasing the overall value created within the business ecosystem.

As mentioned before, data and data management play a crucial role in creating value within the business ecosystem – no matter which innovative technology or method is applied. In the following section, we highlight differentiating capabilities that enable organizations to utilize data and relationships as a second value-driving factor in business ecosystems. With that, we also suggest a new and holistic operating model for business ecosystems.

## 4. TOWARDS AN ECOSYSTEM OPERATING MODEL BLUEPRINT

The transformation of organizations toward business ecosystems is driven by society, customers, and IT innovations. For guiding transformation endeavors, we here provide an operating model blueprint for business ecosystems based on capabilities that are necessary for an effective construction. A graphical representation is shown in Figure 1.

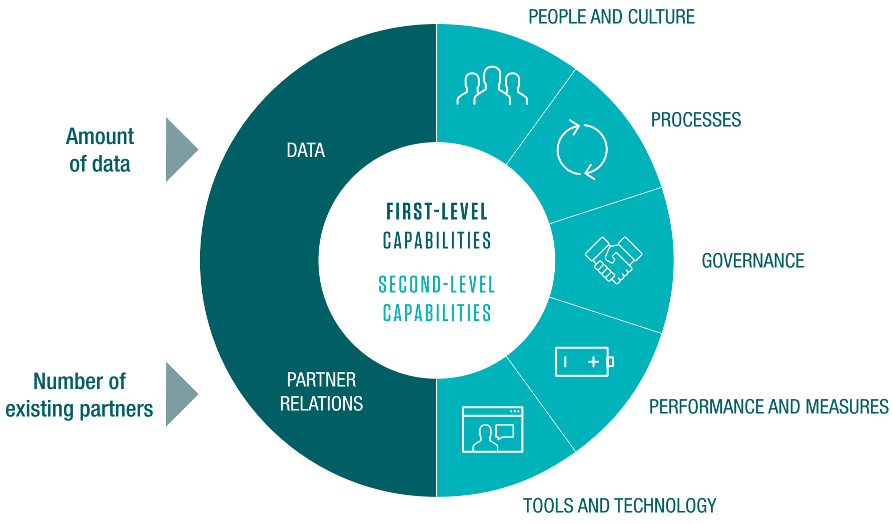
The operating model is constructed of two classes of capabilities: first-level capabilities provide the necessary capabilities for operating a business ecosystem in exploiting the basic resources of such constructs, namely, amount of data and number of existing partners. We define second-level capabilities as those that enable organizations to effectively take part in business ecosystems.

In the following, we elaborate on the two first-level capabilities necessary for operating in business ecosystems, data and partner relations, that ground themselves in the existence of two core resources for business ecosystems – amount of data and number of existing partners, which were already highlighted above. Furthermore, we elaborate on five classes of (second-level) capabilities that enable an organization to take part in a business ecosystem. While the capabilities "data" and "partner relations" provide the essential basic elements for a business ecosystem to create value and are crucial capabilities for its orchestration, the five additional classes provide a home for capabilities necessary to leverage and coordinate the essential basic elements. We base our analysis on the idea that an organization, in general, needs the following elements to function: a "human" element, since organizations consist of humans (people and culture); an operational workflow element, since the organization works in processes (processes); a structural element, since policies and decision-making mechanisms need to be applied (governance); a controlling element, since organizations strive for improvements that are enabled by monitoring and assessing (performance and measures); and a technological element, since organizations today are empowered by technology (tools and technology). We here aim to connect these elements to business ecosystems.

### 4.1 Data

We have already highlighted the importance of data to business ecosystems. However, it is not only the amount of data, but the consistency and possibility to jointly exploit the data within the business ecosystem that is necessary to consider. Besides managing the data within the organization, managing the data across the business ecosystem becomes an important pillar for providing the most value to customers and possibly enable the exploitation of new business opportunities. Imagine, in the case of the Helvetia Home example, the user would be forced to re-enter the same data for each organization taking part within the business ecosystem. In such a setting, the perceived value of the business ecosystem would be significantly reduced. In addition, data is key for recommending the right increments to be connected within the business ecosystem

**Figure 1:** Operating model blueprint for business ecosystems



for a customer-centered solution. Exemplarily, it is necessary to provide information about the type of housing to identify the right solutions for the management of the accommodation – recommending a gardener to an owner whose house does not have a garden does not offer much value for the customer and does not create trust in the solution.

Moreover, managing the data across organizations becomes key beyond the aforementioned examples. Organizations might already have some information on the same customer. Bringing together the data sources within a business ecosystem offers the possibility of increasing data quality by filling gaps in a central data source. This further benefits the application of intelligent methods for identifying patterns and possible new business opportunities.

Different data management capabilities can be considered important for organizations in the realm of business ecosystems. For example, Vetterling and Hackl (2022) focus on data management capabilities and highlight various findings. In the early phases of business ecosystems, capabilities that enable the organization to make decisions based on the available data are particularly relevant. In later-phase business ecosystems, capabilities such as maintaining the highest possible data quality in connection with data can be highlighted as particularly important.

### 4.2 Partner relations

The relationship management between the organizations within the business ecosystem is an important resource, and its orchestration is one of the distinguishing capabilities that differentiates business ecosystems from other value-creation settings. Each organization entering the business ecosystem brings its own partner relations to the overall setting, thereby increasing the number of connections within the overall business ecosystem. However, it is not only the number of partners that counts but the underlying logic that an organization cannot act alone in a business ecosystem. Organizations are dependent on each other's increments for the overall service to be combined. The complementarity of the increments is considered one of the basic elements in business ecosystems and distinguishing them from other networked value creation settings [Jacobides et al. (2018)]. The higher the degree of complementarity, the higher the possible value created when the increments are correctly combined in one service offering. However, a high degree of complementarity also brings risks, as the failure of one organization to provide the right increment threatens the entire business ecosystem [Lingens et al. (2023)].

For organizations, it is, therefore, necessary to shift away from seeing partners as sole providers of sub-products for the good of a product offering provided by the organization towards considering them as partners in a joint endeavor of providing as complete a service as possible. Organizations need to work together on equal terms with the other organizations in the ecosystem.

Shifting the focus away from delimitation towards coopetition was already mentioned above. This includes not only working in a network of organizations but also identifying "the right" organizations to work with. Based on an example of the housing sector, Maicher et al. (2023) highlight the importance of identifying and orchestrating shared values between partners. Furthermore, an element of trusting each other needs to be mentioned here. Data was already mentioned as a key resource of business ecosystems, and we also highlighted not only the necessity but also the benefits of sharing data. Nevertheless, sharing the resulting information between organizations demands that it is not excavated for a one-sided benefit.

What further complicates the environment is that partners in business ecosystems might change over time despite all efforts to sustain the relationship. New partners then need to be identified, providing the necessary increments for the overall service offering, and need to be integrated into the group of existing actors.

Data and partner relations form the first level capabilities that are necessary for value creation in business ecosystems. In the following section, we present the second-level capabilities that are necessary for organizations to effectively take part in business ecosystems.

### 4.3 People and culture

Business ecosystems pose a challenge to both organizational structures and individual employees, requiring a shift from an isolated to a collaborative mindset.

Maicher et al. (2023) emphasize the importance of a clear organizational mindset and internal clarity about roles within the ecosystem and advocate for a culture that celebrates shared success and supports adaptability in dynamic environments. This adaptability is critical because business ecosystems are inherently dynamic and require a resilient organizational culture to manage change and uncertainty.

It helps organizations to learn from the past and to be able to experiment to discover new business opportunities. Such organizational capabilities as network learning [Buck et al. (2021)], experimentation with business opportunities, and exploiting these [(Achtenhagen et al. (2013)] are already beneficial for organizations today and have an amplified power in business ecosystems.

Lastly, business ecosystems are networks of organizations. Hence, collaboration is a key aspect within and across organizations. Regarding ecosystems, Schreieck et al. (2021) highlight the "digital business innovation capability" as a capability of organizations to drive innovation through a collaboration of independent actors. This capability might be grounded in the people who form the organization, since they need to have the mindset to collaborate with other organizations. It might even lead to joining forces with competitors to create a valuable service offering. Such a setting can be observed in the business ecosystem of Well Gesundheit AG [Well Gesundheit AG (2023)] in Switzerland, where the healthcare insurers Visana and CSS, normally competitors, joined forces to innovate in the Swiss healthcare sector.

## 4.4 Processes

The consequent establishment of coordinated and scalable processes is the necessary basis for a successful business ecosystem. In general, processes define how activities are organized, managed, and executed. In a business ecosystem, these processes do not end at one organization's borders but span organizational borders within the business ecosystem. In addition, due to the dynamic element in business ecosystems, processes must be easily adaptable for change.

Already today, reconfiguring processes toward a specific target is an essential organizational capability [Buck et al. (2021)]. This capability is even more critical in business ecosystems when considering the dynamic structure of the overall system [Moore (1993)]. Further necessary organizational capabilities regarding processes might be seen in establishing and sustaining connections between different entities – within and across organizations [Buck et al. (2021)]. This capability might be further amplified in its importance in business ecosystems due to their networked structure. In addition, the automatization of processes by applying state-of-the-art technology might help to pave the way toward scalability of the process stack.

## 4.5 Governance

Providing the right governance framework for a business ecosystem is one of the pressing challenges in establishing and maintaining such a construct [Pidun et al. (2020), Schaefer et al. (2023)]. Overall, the governance needs to ensure that all partners are working towards a shared goal and underline the "on-equal-terms" working mode of all partners.

Regarding business ecosystems, organizations need to distribute power across the partners in order to establish a basis for jointly developing the system. This can be done in several ways, exemplarily by having a consortium building the backbone for the development or by providing a shared ownership structure. A good example for the first exemplary type might be observable in myky, the business ecosystem in the housing sector in Switzerland. Here, a consortium of banks, as well as an insurance company and an energy provider are backing the development of the business ecosystem. An even more powerful approach for distributing power is undertaken by SmartWE – who are building a CRM platform that is set up as a self-owned participatory joint stock company [(SmartWE World SE (2023)]. In essence, this means that participants in the business ecosystem around the platform can acquire shares of the stock company, through which they receive corresponding voting rights. As is usual with public limited companies, the voting rights regulate the co-determination. If a participant leaves the business ecosystem, they must return their shares. In addition, there are limits to the maximum number of shares that can be acquired by one organization, to avoid a corresponding concentration.

To optimize the utilization of data and partner relations within a business ecosystem through technology, it is necessary to implement mechanisms for monitoring activities and identifying wrongdoing. In addition, having punishment mechanisms in place for identified wrongdoing is important. If a group of partners are working together on the same foundational elements, in our case it is data and partner relations, then mechanisms need to be established that enable the partner to behave in a way that benefits all and not just some. In order to establish such governance mechanisms, the use of technology, among other things, may offer added value. For example, Schaefer et al. (2023) point to the prospects of blockchain technology to help solve governance problems in business ecosystems.

## 4.6 Performance and measures

In an earlier publication, we had proposed a set of performance measures that could generally enable the controlling of business ecosystems [Vetterling and Baumöl (2023)]. Here, we want to focus on aspects related to exploiting the grounding resources of business ecosystems, "amount of data" and "number of existing partners". Consequently, we identify here three areas of development for capabilities based on the different levels of value creation. These areas are customer interaction, the backend for service creation, where the activities of the individual organizations are orchestrated, and the overall business ecosystem [Vetterling and Baumöl (2023)].

Considering the customer interaction, the performance of AI-based applications, such as chatbots, can be measured by considering successful customer interactions, since not only the successful provision of the right service needs to be taken into account, but also the perception of the interaction by the customer. The generated data then needs to be used in a learning loop to further improve the service offered.

In the backend, especially in later stage business ecosystems, efficiency needs to grow. Hence, inefficiencies can be identified by analyzing the data using intelligent methods. Furthermore, such methods can be used to predict possible bottlenecks when the whole system grows.

Organizations in business ecosystems need to not only monitor their own organizational performance but also the overall performance of the entire business ecosystem. This shift in perspective, as well as the respective controlling instruments, need to be established in the individual organizations in order to successfully engage in business ecosystems.

## 4.7 Tools and technology

The use of technology is promising in the context of business ecosystems. Business ecosystems are more complex than other value networks. Accordingly, intelligent methods can ensure efficient control on a scaled basis. Analyzing and evaluating large amounts of data is one of the core applications where new technologies can be utilized in business ecosystems.

Image generated by Adobe Firefly

Furthermore, the provision of (technical) interfaces, such as application programming interfaces (APIs), is a necessary prerequisite for the cooperation of different organizations in the business ecosystem. These interfaces facilitate seamless interaction and data exchange, thereby bolstering the interconnectedness and cooperative synergy essential for the thriving of business ecosystems.

To create and put IT solutions into action in a manner that meets business requirements in a way that is both cost-effective and well-managed is an essential organizational capability [Wang et al. (2012)]. This capability might be particularly important for operating models in business ecosystems. Furthermore, as a mixture of different capabilities might be needed to be implemented to generate the best possible tech stack, the organizational capability to use different technologies, which Buck et al. (2021) identified as technological capability, might be a differentiating factor in business ecosystems.

## 5. CONCLUSION

Innovative technologies and applying intelligent methods capable of performing tasks that on a scaled basis enable new value propositions and support new organizational designs for value creation are today's differentiating factors. Based on the example of business ecosystems, it becomes clear how these innovations impact value creation. New service offerings can be promoted for the customers, and new ways of creating these service offerings are enabled as well. This leads to adapting the existing operating models by focusing on data and partner relations as a main lever. Hence, capabilities for managing these need to be developed with priority. The second-level capabilities enable goal-oriented part-taking in the business ecosystem from the individual company's perspective.

For the future of developing business ecosystems and – as a special instance – financial business ecosystems, both the structural and incentive-driven organizational design have to be transformed. Data and partner relations are understood as important drivers and with that collaboration and standardization capabilities must be developed. That also requires a transformation of the mindset in the individual organization to form the mindset for the business ecosystem. This journey must be carefully planned and coached in a joined effort.

## REFERENCES


Achtenhagen, L., L. Melin, and L. Naldi, 2013, "Dynamics of business models - strategizing, critical capabilities and activities for sustained value creation," Long Range Planning 46:6, 427-442

Adner, R., 2017, "Ecosystem as structure: an actionable construct for strategy," Journal of Management 43:1, 39-58

Buck, C., T. Grüneke, and K. Stelzl, 2021, "Structuring the jungle of capabilities fostering digital innovation," in Ahlemann, F., R. Schütte, and S. Stieglitz (eds.), Innovation through information systems, Springer International Publishing.

Dattee, B. A., L. D. W. Thomas, and E. Autio, 2023, "The limits of innovation ecosystem strategy: a demand-side view," Academy of Management Proceedings 2023:1

Dumitru, V., D. Iorga, S. Ruseti, and M. Dascalu, 2023, "Garbage in, garbage out: an analysis of HTML text extractors and their impact on NLP performance," Proceedings - 2023 24th International Conference on Control Systems and Computer Science, CSCS 2023, 403-410

Helvetia. (n.d.)., 2023, "Vision of the home ecosystem," in Seehofer, T., and C. Lechner, The evolution of the "home" ecosystem: the case of Helvetia, University of St. Gallen, 11, https://tinyurl.com/e9a2y9uz

Jacobides, M. G., C. Cennamo, and A. Gawer, 2018, "Towards a theory of ecosystems," Strategic Management Journal 39:8, 2255-2276

Lingens, B., V. Seeholzer, and O. Gassmann, 2023, "Journey to the Big Bang: how firms define new value propositions in emerging ecosystems," Journal of Engineering and Technology Management, 69(July), https://tinyurl.com/mr25cxy2

Maicher, L., T. Lenoci, and D. Vetterling, 2023, "Controlling-erkenntnisse beim Aufbau eines business ecosystems," Controlling - Zeitschrift Für Erfolgsorientierte Unternehmenssteuerung 05, 33–38

Moore, J. F., 1993, "Predators and prey: a new ecology of competition," Harvard Business Review 71:3, 75-86

Moore, J. F., 2023, "Navigating the death of competition: the emergence of business ecosystems and beyond," Network Law Review, https://tinyurl.com/2p8nywkn

myky AG., 2023, myky. https://www.myky.ch/

Pidun, U., M. Reeves, and M. Schüssler, 2020, "Why do most business ecosystems fail?" Boston Consulting Group, June 20, https://tinyurl.com/45uaz48d

Schaefer, B., D. Vetterling, L. Wrobel, and R. Jung, 2023, "Where blockchain can help – an analysis of challenges in ecosystems blockchain technology may help to solve," AMCIS 2023 Proceedings 24, 1-5, https://tinyurl.com/bdcvda34

Schreieck, M., M. Wiesche, and H. Krcmar, 2021, "Capabilities for value co-creation and value capture in emergent platform ecosystems: a longitudinal case study of SAP's cloud platform," Journal of Information Technology 36:4, 365-390

Seehofer, T. M., and C. Lechner, 2023, "Evolution of the "home" ecosystem: the case of Helvetia," The Case Centre, University of St. Gallen, https://tinyurl.com/57avvm9s

SmartWE World SE, 2023, SmartWe – Wir, https://tinyurl.com/tdyjkzu2

Tansley, A. G., 1935, "The use and abuse of vegetational concepts and terms," Ecology 16:3, 284-307

Vetterling, D., and U. Baumöl, 2023, "Ecosystems als Quelle kundenzentrierter Wertschöpfung," Controlling - Zeitschrift Für Erfolgsorientierte Unternehmenssteuerung 05, 04-11

Vetterling, D., and T. Hackl, 2022, "Data management im Ecosystem-lebenszyklus," Controlling - Zeitschrift Für Erfolgsorientierte Unternehmenssteuerung 34, 12-19

Wang, N., H. Liang, W. Zhong, Y. Xue, and J. Xiao, 2012, "Resource structuring or capability building? An empirical study of the business value of information technology," Journal of Management Information Systems 29:2, 325-367

Well Gesundheit AG., 2023, Über uns. Über Uns, https://tinyurl.com/56ue8eck

Yao, L., A. Cahill, and D. F. Mccaffrey, 2020, "The impact of training data quality on automated content scoring performance," Association for the Advancement of Artificial Intelligence

Zukunftsinstitut GmbH, 2023, Die Megatrends, https://tinyurl.com/2phstcsf

# DEVELOPERS 3.0: INTEGRATION OF GENERATIVE AI IN SOFTWARE DEVELOPMENT

**FAYSSAL MERIMI** | Managing Principal, Capco
**JULIEN KOKOCINSKI** | Partner, Capco

## ABSTRACT

The concept of "developers 3.0" is emerging, defining the new avant-garde generation of software development professionals. These developers, transcending traditional skills, place generative artificial intelligence (AI) at the heart of their approach, thus revolutionizing software design and development paradigms. This article explores the methodologies and strategies adopted by these innovators, highlighting notable advantages in terms of productivity and quality. At the same time, we address the challenges associated with this combination of traditional software development practices with the new methodologies centered around generative artificial intelligence, such as ethical issues, security concerns, and the need to maintain a balance with traditional skills. Our analysis aims to provide an in-depth perspective on the growing influence of generative AI in the field of software development and its implications for the future of the profession.

## 1. INTRODUCTION

The history of software development is constantly shaped by revolutionary technological innovations. Today, generative artificial intelligence (AI) stands out as one of the most significant advancements, opening the door to a myriad of questions and possibilities. What are the different types of generative AI and how do they position themselves in the current technological tools landscape? In the face of this rise, does generative AI represent a threat to developers or an opportunity? It is in this context that we introduce the concept of "developers 3.0", innovators who skillfully navigate between traditional skills and advanced AI capabilities. In this article, we will explore the role of generative AI in software development, its benefits, its challenges, and how it shapes the future of the profession.

## 2. WHAT ARE THE DIFFERENT TYPES OF GENERATIVE AI?

There are two major categories of generative artificial intelligence. The first category encompasses universal generative AIs. These systems are trained on a very large public dataset and aim to address a variety of questions or queries, much like an online search engine. The second category pertains to organization-specific generative AIs. These solutions are designed to access only the data of a specific entity. Typically, these models are also specifically trained to handle particular tasks, such as customer relationship management or the synthesis of legal documents.

In the first scenario, every interaction and feedback from users contribute to the improvement of the underlying algorithm of the generative AI. In the second scenario, interactions only benefit the private version of the organization's own AI, which can be hosted by the provider or by the organization itself.

The market is currently flooded with AI-based tools, but not all are created equal. It is crucial to choose tools that fit one's needs. These tools, spread across various application domains, form a dynamic and constantly evolving landscape, illustrating the depth and diversity of AI-based solutions available to developers. Figure 1 provides an overview of this landscape, highlighting the main categories of tools and how they interact with each other.

Table 1 provides a non-exhaustive list of tools specific to code generation and software development.

**Table 1:** Tools specific to code generation and software development

| CODE GENERATION | |
|---|---|
| **Application** | **Description** |
| aiXcoder | AI-based code completion tool |
| Bito | Assists developers in code generation |
| CodeAssist | Offers real-time code suggestions |
| CodeComplete | Automates code generation for common tasks |
| CodeGPT | AI model for code generation |
| Codel | Facilitates code writing with AI-based suggestions |
| Codeium | Intelligent code completion tool |
| GitHub Copilot | Coding assistant powered by OpenAI |
| CodeGuru | Amazon tool for AI-based suggestions |
| Mutable | Code generator based on project needs |
| Replit Ghostwriter | Offers code suggestions while writing |
| Tabnine | AI-based code completion for various IDEs |
| Warp AI | Speeds up the coding process with automated suggestions |
| **CODE ANALYSIS AND DEVOPS** | |
| **Application** | **Description** |
| Adrenaline | Analyzes code to optimize performance |
| AI Code Reviewer | Examines code for potential errors |
| Codacy | Code quality analysis platform |
| Codeball | Analyzes code to improve quality |
| Coderbuds | Collaborative code analysis tool |
| Codiga | Checks code quality in real-time |
| Metabob | Analyzes code to detect bugs |
| What The Diff | Compares code versions to detect changes |
| Whispr | AI-based code analysis tool |

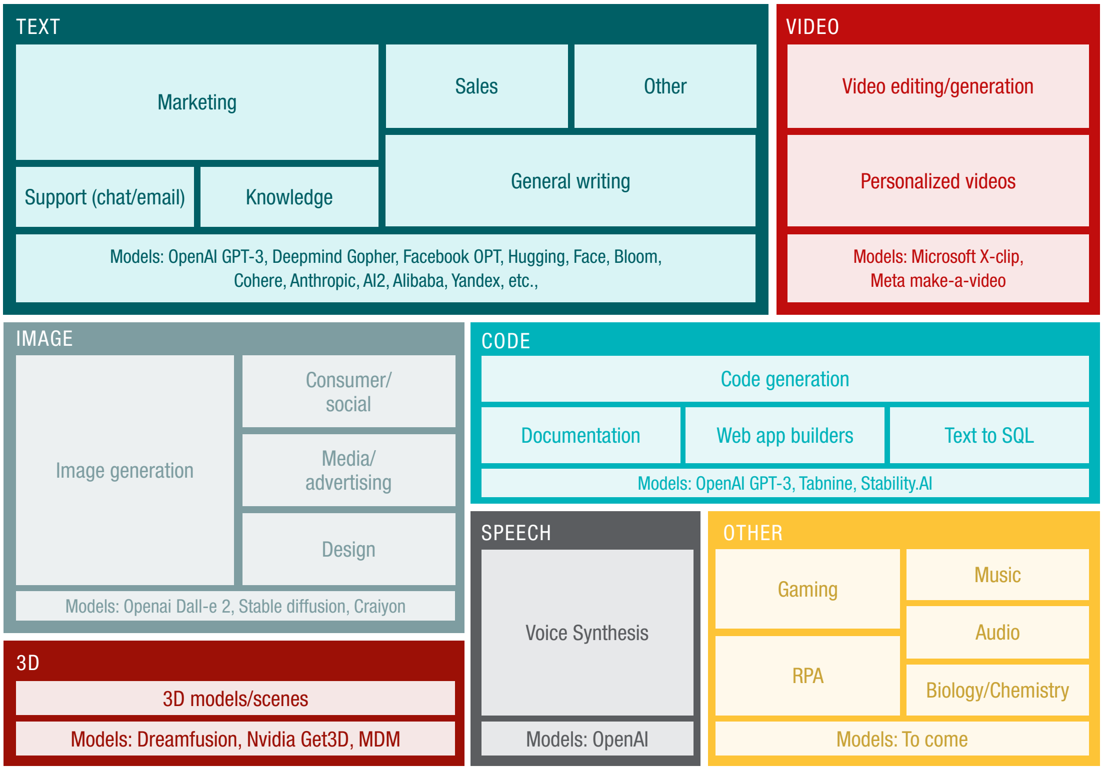
| DOCUMENTATION GENERATION | |
|---|---|
| **Application** | **Description** |
| DocumentationLab | Automatically generates documentation from code |
| DocuWriter | AI-based documentation writing tool |
| FigStack | Creates technical documents with AI-based suggestions |
| Mintlify | Transforms code into readable documentation |
| Stenography | Automated documentation tool for projects |
| **DATA** | |
| **Application** | **Description** |
| AI2sql | Converts natural language questions into SQL queries |
| Channel | Analyzes data to provide insights |
| Chat2Stats | Transforms conversations into statistics |
| Consensus | Collaborative data analysis tool |
| Dataherald | Generates reports based on data analysis |
| Defog | Clarifies data for better understanding |
| GenerativeBI | AI-based data analysis tool |
| Finalle | Automated data analysis platform |
| Kanaries | Transforms data into visualizations |
| Lookup | AI-based data search tool |
| Maya | Assists in data analysis with AI-based suggestions |
| ProbeAI | Real-time data analysis platform |
| SQL Genie | Assists in creating SQL queries |
| SQL Genius | Suggests optimizations for SQL queries |
| String | Textual data analysis tool |
| Symbl | Converts conversations into analyzable data |
| TableTalk | Transforms data into natural language speech |
| Windsor | Visual data analysis platform |

## 3. IS GENERATIVE AI A THREAT TO DEVELOPERS?

Generative AI, although a powerful and constantly evolving technology, is not necessarily a threat to software developers. In fact, there are reasons to believe that generative AI can be viewed as an asset, rather than a threat, to developers:

- **Complementarity rather than substitution:** generative AI is designed to complement developers' skills, not replace them. It can automate certain repetitive tasks, but human creativity, logic, and expertise remain indispensable for many aspects of software development.
- **Complexity of real projects:** while AI can generate code for specific tasks, the complexity, architecture, and business logic of real projects require deep understanding and human expertise.
- **Validation of AI-generated code:** regardless of the source of the code, whether generated by AI or written manually, rigorous control is essential. Developers must always check, test, and validate the code to ensure its quality and functionality.

**Figure 1:** The generative AI application landscape




Source: Derived from Sequoia Capital[1]

- **Human interactions:** software development is not just about code. It is also about understanding client needs, working as a team, communicating ideas, and solving problems together. These interpersonal skills cannot be replaced by a machine.
- **Ethical and moral boundaries:** there are decisions in software development that require ethical and moral judgment, such as considering user privacy or creating socially impactful applications. These decisions require human thought.
- **Adaptability:** needs and technologies are constantly evolving. Developers can adapt, learn, and change direction based on changing requirements, a flexibility that AI has not yet achieved.

According to a recent survey conducted by Stack Overflow, which asked the opinions of over 90,000 developers regarding the adoption of generative AI,[2] AI is already firmly rooted in the daily lives of developers. The study finds that 44% of them actively use AI-based tools in their processes, and another quarter plan to follow this trend soon. It is particularly notable that 55% of programming novices turn to these tools from the start of their learning.

However, the issue of trust remains concerning. Only 3% of the developers surveyed have absolute trust in these AI tools for development. Conversely, 6% express total reservations. The majority, or 39%, are cautiously optimistic, falling into a zone of moderate trust.

The study also highlighted regional variations in AI adoption. Developers based in India, Brazil, and Poland are more inclined to integrate these tools, while their counterparts in the European Union and the United States are more reserved.

Looking to the future, the transformative impact of AI on software development is undeniable. An impressive majority, 77% of the developers surveyed, anticipate that AI will redefine their way of writing and debugging code in the near future. These results suggest an imminent shift in the software development landscape, with generative AI as a major protagonist.

1 https://tinyurl.com/yvwrjpfx
2 https://tinyurl.com/mtsvfktj

## 4. WHO ARE "DEVELOPERS 3.0" AND HOW DO THEY APPLY GENERATIVE AI?

Developers 3.0 are not just coders; they are innovators. They understand that AI is not a threat but a tool that can be used to augment their capabilities. They are curious, always ready to learn and adapt, and see AI as a collaborative partner rather than a replacement. For instance, take the case of GitHub Copilot, an AI-based tool that suggests lines of code as developers write. Instead of seeing this as a threat, many developers have embraced it as an assistant that speeds up their coding process. Similarly, companies like DeepMind have used AI to optimize energy consumption in data centers, assisting engineers in identifying more efficient solutions. In the realm of game design, AI is used to generate levels or scenarios, allowing developers to focus on other creative aspects of the game. These examples demonstrate how Developers 3.0 integrate AI into their daily work, not as a rival, but as a valuable collaborator.

### 4.1 Generative AI in software development

Generative AI in software development is transforming the way developers approach and manage their projects. Here, we present a detailed overview of its capabilities and applications:

- **Code generation:** generative AI can produce code from a natural language description, allowing developers to quickly translate their ideas into functional code.
- **Optimization:** it analyzes and optimizes the code to enhance performance, reduce redundancy, and ensure the software operates optimally.
- **Error detection and correction:** AI predicts and identifies potential errors, offering a proactive form of debugging. It also suggests corrections, reducing debugging time.
- **Recommendations:** AI suggests relevant methods, libraries, or approaches for the project, helping developers stay updated with best practices.
- **Automated testing:** it generates and runs tests based on the code, ensuring comprehensive coverage and software robustness.
- **Documentation:** generative AI can automatically generate relevant documentation based on the code, facilitating project understanding and maintenance.
- **Language transformation:** it is capable of translating code from one programming language to another, easing portability and integration across different platforms.
- **Code explanation:** AI can provide detailed explanations about how the code operates, assisting developers in understanding complex, or automatically generated, code segments.
- **Code completion:** by suggesting real-time code segments or structures, AI aids developers in coding more swiftly and efficiently.
- **Simplification suggestions:** AI analyzes the code to pinpoint segments that can be simplified or refactored, ensuring cleaner and maintainable code.

By integrating these capabilities, generative AI offers developers a powerful suite of tools that not only enhance the quality of work but also expedite the development process.

### 4.2 Illustration of the use of generative AI in software development

Generative AI, with its advanced capabilities, is revolutionizing the way we approach software development. To concretely illustrate its impact, we refer to a recent article written by our colleagues at Capco,[3] which assessed the reliability of AI tools in the software development process using a real-life case of a financial institution. The primary goal of this study was to demonstrate how generative AI can transform technological operational models throughout the software delivery process.

The study focused on six common activities of the software delivery process, including requirement writing, architecture, design, user experience, code writing, testing, and DevOps. Tools developed by OpenAI were used to generate assets typically created during the delivery phase. These assets were then qualitatively evaluated on three dimensions: product quality, time savings, and resource/cost savings.

The results showed that using generative AI for these activities led to significant time savings, especially for code writing. Even in areas where AI's performance was lower, the products generated by AI were useful for verifying the completeness of human-generated assets.

The article concludes by highlighting the potential benefits of using generative AI in the software delivery process, including time savings, improved quality, and cost reduction. The authors recommend that CIOs adopt a systematic approach to integrating AI into the software delivery process, invest in training the necessary skills, and consider using local versions of AI tools to minimize risks associated with using cloud-based versions.

[3] https://tinyurl.com/2fmd569z

## 5. BENEFITS AND CHALLENGES OF INTEGRATING AI

### 5.1.Benefits

It is our opinion, and suggestion in this article, that integrating AI within software development has a number of advantages. Some of these advantages are highlighted below.

- **Increased productivity:**
  - **Task automation:** repetitive tasks, such as generating code for common functions or detecting common errors, can be automated. This frees up time for developers, allowing them to focus on more innovative aspects of development.
  - **Debugging time reduction:** with AI tools that quickly identify errors, the time spent debugging is significantly reduced.
- **Quality enhancement:**
  - **In-depth analysis:** AI can scan thousands of lines of code in seconds, identifying errors or inefficiencies that the human eye might miss.
  - **Real-time optimization:** some AI tools can suggest optimizations in real-time, as code is being written, ensuring optimal performance from the outset.
- **Continuous learning:**
  - **Proactive suggestions:** AI can suggest new methods or techniques based on current industry trends, helping developers stay up to date.
  - **Integrated training:** with tools like GitHub Copilot, developers can receive AI-based suggestions as they code, offering a real-time learning opportunity.
- **Enhanced collaboration:**
  - **AI-assisted code review:** AI tools can assist in code review by suggesting improvements or identifying potential issues, facilitating collaboration among team members.
  - **Project management:** AI can also assist in project management by predicting delivery timelines, identifying bottlenecks, and suggesting resource reallocations.
- **Customization and adaptability:**
  - **Coding style adaptation:** some AI tools can adapt to a developer's specific coding style, offering suggestions that match their personal preferences and practices.
  - **Seamless integration:** AI-based tools are designed to integrate seamlessly into existing development environments, offering benefits without disrupting established workflows.

### 5.2 Limitations and challenges

As with most things in life, nothing comes without challenges, and generative AI is no exception. Below, we highlight some of the challenges that developers face when trying to integrate AI within their systems:

- **Skill loss:**
  - **Skill atrophy:** if developers rely too heavily on AI for routine tasks, they risk losing practice and mastery of certain fundamental skills.
  - **Less hands-on training:** with AI taking over complex tasks, new developers might miss out on essential learning opportunities.
- **Complexity:**
  - **Obfuscated code:** code generated by AI can sometimes be verbose or structured in a way that is not intuitive for humans.
  - **Tool dependency:** if an AI tool becomes obsolete or is no longer supported, it could pose issues for projects heavily dependent on its generated code.
- **Ethics:**
  - **Accountability:** in case of a failure or error caused by AI-generated code, determining accountability can be tricky. Is it the fault of the tool, the developer who used it, or the organization that adopted it?
  - **Transparency:** decisions made by AI are not always transparent, which can pose ethical challenges, especially in sensitive areas like healthcare or finance.
- **Security:**
  - **Potential vulnerabilities:** automatically generated code might introduce unintentional vulnerabilities, exposing applications to risks.
  - **Dependency on external sources:** if AI relies on external data or libraries to generate code, it could introduce unexpected security risks.
- **Cost:**
  - **Initial investment:** adopting AI tools for development might require a significant initial investment in terms of training, licensing, or integration.

- **Vendor dependency:** once an organization commits to a specific AI tool, they might become dependent on that vendor, potentially leading to additional long-term costs.
- **Interpretability:**
  - **Lack of justification:** AI might suggest or generate code without providing clear justification, making it challenging for developers to understand or justify certain coding decisions.

## 6. PERSPECTIVES AND THE FUTURE

As generative AI continues to evolve, developers 3.0 will find themselves at the crossroads between tradition and innovation.

This provides new perspectives on the future of these avant-garde developers, who will need to constantly update their skills, to understand the nuances of generative AI, as well as learn how to collaborate effectively with it. Indeed, generative AI will not just be a tool but a collaborator. Developers 3.0 will work in tandem with AI systems, where AI might suggest solutions, optimize code, or even anticipate the developer's needs.

Continuous learning will become the norm for developers 3.0, with an increased emphasis on understanding AI systems and integrating them into software solutions. For example, generative AI will allow for unprecedented personalization of software solutions. Developers 3.0 will then be able to create applications that adapt in real-time to users' needs, offering highly personalized user experiences at speed.

With generative AI increasingly making autonomous decisions, questions of ethics and responsibility will come to the forefront. Developers 3.0 will need to navigate this complex landscape, determining who is accountable for errors or issues caused by AI-generated code.

Impact for organizations will be massive. On the one hand, there will be challenges as adoption of generative AI increases, requiring new roles, such as "AI collaboration engineer" or "AI ethics specialist", reflecting the changing nature of software development. This will generate a deep cultural change, requiring supporting the organization at all levels. On the other hand, there will be many new opportunities. As an example, by combining crowdsourcing and AI, organizations will pave the way for more advanced and efficient software engineering methods. Leveraging the collective wisdom of a multitude and complementing it with AI skills will bring new benefits, such as accelerated problem solving and higher-quality outputs, and stimulate innovation in software design.

> *"Developers 3.0 integrate AI into their daily work, not as a rival, but as a valuable collaborator."*

## 7. IMPACT ON THE FINANCIAL SERVICES INDUSTRY

Generative AI is expected to have a significant impact on the financial sector, and developers 3.0 will play a key role in transforming this sector, creating solutions that fully harness the potential of AI.

Financial services, known for stringent regulations and cautious approach to data – be it personal, client, or banking data – demands specialized AI solutions. Most immediate use cases, and easier in terms of compliance and legal validation, would revolve around organization-specific generative AIs. These solutions are designed to access internal data, with the possibility of accessing external data as well, but with the guarantee that no sensitive information is transmitted outside of the institution. Such solutions could address existing pain points while allowing institutions to experiment with, and learn from, generative AI.

For instance, knowledge and document management can greatly benefit from this technology, enabling all departments to swiftly obtain answers based on internal intelligence and experience.[4] Within IT application maintenance teams, IT developers and analysts powered by generative AI will become more effective in solving bugs and issues, leveraging past experience and data.

Regulatory compliance is a constantly evolving challenge in the financial services sector. Generative AI can assist banks in this area as well by generating checklists and updates aligned with the latest regulatory directives. It can also bolster fraud detection by generating scenarios based on known fraud patterns, thereby identifying potential system vulnerabilities, and proposing new solutions, rules, and strategies.

[4] https://tinyurl.com/34zmcdnb

Image generated by Adobe Firefly

Customers interactions will significantly evolve thanks to generative AI. Relationship managers, call centers, claim management teams, or anyone having interactions with customers and collaborating with generative AI will be able to improve customer experience and satisfaction, through quicker, more precise and tailored communication and propositions. Generative AI will also power more sophisticated consumer-facing chatbots.

Moreover, by anticipating market and consumer trends, generative AI can accelerate the creation of innovative financial products tailored to consumers' evolving needs.

Developers, especially developers 3.0, are pivotal in this transformative journey. The future of banking and insurance envisions adaptive systems. These systems, powered by generative AI, can realign in real-time to economic shifts, ensuring efficient asset and liability management.

The advent of universal generative AIs in software development offers financial institutions a unique opportunity to modernize legacy systems cheaper and faster. Developers 3.0 can interpret current code, even in languages that might have become obscure, draft specifications based on this code, hasten adaptations with the users, and transition this code to newer technologies at speed.

With the world's markets becoming increasingly interconnected, developers 3.0 will leverage generative AI to devise highly secured tools and solutions that preemptively manage risks associated with global fluctuations and threats.

In the evolving landscape of the financial services sector, the expertise of developers 3.0 combined with the capabilities of generative AI promises a future marked by efficiency, security, and personalization. As financial institutions steer through the complexities of modernization and regulatory adherence, generative AI emerges as a symbol of innovation and forward momentum.

## 8. CONCLUSION

The advent of generative AI marks a pivotal milestone in the evolution of software development. The concept of "developers 3.0", which we introduced in this article, epitomizes this new breed of professionals, who, equipped with traditional skills and a profound understanding of AI, are poised to redefine the boundaries of what is achievable in the software development realm.

These trailblazing developers are not merely technicians, they are visionaries who recognize the potential of generative AI as a collaborative tool, rather than a threat. They stand at the nexus of human-machine synergy, working hand-in-hand with AI systems to craft solutions that are more innovative, efficient, and tailored.

However, as we have discussed, this evolution does not come without its challenges. Ethical dilemmas, accountability, and the need for ongoing education and cultural change arise, demanding thoughtful and coordinated efforts from the developer community, businesses, and policymakers.

Ultimately, the future of software development is bright and brimming with uncharted possibilities. The developers 3.0 are at the forefront of this new era, and we eagerly anticipate the innovations and transformations they will bring to industry and beyond.

# DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE IN ORGANIZATIONS

**NIRAN SUBRAMANIAM** | Associate Professor in Financial Management & Systems, Henley Business School

## ABSTRACT

Digital transformation revolutionizes how businesses provide value by seamlessly integrating digital technologies into operations, strategies, and culture. Its core objectives encompass enhanced efficiency, elevated customer experiences, and heightened competitiveness, while ensuring adaptability in the face of swiftly evolving technology and market landscapes. A key enabler in this transformation is artificial intelligence (AI), which infuses intelligence and automation into digital technology utilization. AI's capabilities encompass mining and analyzing diverse organizational data to unearth patterns that drive recommendations and inferences. For instance, customer data analysis unveils preferences, enabling personalized marketing and lucrative opportunities such as cross-selling and up-selling. AI, with its pattern recognition, inference, recommendation, and predictive analytics, is at the forefront of driving digital transformation in organizations. This article proposes a framework for successful digital transformation in organizations.

## 1. INTRODUCTION

"Digital" and "transformation" are perhaps the most frequently used words in management speak today. While transformation means different things to different audiences, its value and significance are neither misunderstood nor underestimated by organizations. Ironically, the Cambridge and Oxford definitions also agree on the meaning of transformation as "a complete change in the appearance or character of something or someone..." The ubiquitous nature of digital technologies since the advent of computing has only made digital transformation an ideal for companies around the world. As various technologies mature and proliferate, the nature and potential of certain technologies, such as AI, take prominence as enabling technologies with ample promise in digitally transforming organizational realities.

### 1.1 Transforming organizations

Mark Twain said, "if you want to change the future, you must change what you are doing in the present." While this may seem obvious, organizations tend to settle into norms and practices to often continue under the pretense of making "continuous improvements" while maintaining the status quo.

Transformation is a complete change – a "new" state that is different to the former. Transforming organizations deliver superior customer experience and greater value to their shareholders by systematically studying and effecting changes to their people, processes, and systems. For instance, by adopting newer tools and technologies to make changes in the way people work together[1] and utilizing enabling technologies to fundamentally change end-to-end business processes to deliver operational efficiencies, organizations can move towards digitally transforming to a new state, a new and better organizational reality.

Studies led by management consultants working in different sectors show that although more than 80% of firms embark on some form of digital transformation, less than 20% realize their intended benefits.[2] This statistic on the low success rate is hard to ignore, as the reality seems all too familiar with emerging technologies of the past – business process reengineering

[1] Subramaniam, N., J. Nandhakumar, and J. Baptista, 2013, "Exploring social network interactions in enterprise systems: the role of virtual co-presence," Information Systems Journal 23:6, 475-499

[2] https://tinyurl.com/ys3xe9a9

**Table 1:** Drivers for, and enablers of, digital transformation

| DRIVERS | ENABLERS |
|---|---|
| **Innovative business model** | Create new revenue streams and market opportunities by reimagining business models to leverage digital channels and technologies. |
| **Agility** | Agile decision making and resource allocations by adapting to changing market dynamics, facilitated by digital tools and platforms. |
| **Customer centricity** | Deliver custom solutions, and improve customer satisfaction, placing the customer at the center of business strategies and using digital tools to enhance customer engagement. |
| **Technology adoption** | Embrace technologies such as cloud computing, data analytics, artificial intelligence, the Internet of Things (IoT), and automation to streamline processes, gain insights, and enhance decision making. |
| **Cultural shift** | Foster a culture of innovation, collaboration, and digital literacy within the organization by encouraging employees to embrace and drive digital initiatives. |
| **Collaboration** | Create synergies with partners, suppliers, and even competitors to unlock new value by collaborating within the industry's digital ecosystems. |
| **Data utilization** | Collect, analyze, and utilize data to understand customer behaviors, optimize operations, and create personalized experiences by leveraging data as a strategic asset. |
| **Security and compliance** | Implement robust cybersecurity measures and ensure regulatory compliance as data and digital processes become more critical. |

(BPR) in the eighties, enterprise resource planning (ERP) systems in the nineties, service oriented architectures (SOA) in the new millennium, and application program interfaces (API) in the last decade. Despite the promise and potential of these technologies, success has been much harder to achieve. Aside from investing more on technologies and digitalization initiatives, what are organizations doing to increase the success rate? What are the strategic imperatives to ensure success in digital transformation? Digital transformation initiatives often stall because of a number of challenges, yet there are opportunities that lead to viable strategies for ensuring successful outcomes.

## 2. WHAT IS DIGITAL TRANSFORMATION?

Digital transformation encompasses a broader strategic shift in thinking to fundamentally change the way in which a business delivers value to customers. By integrating digital technologies into various facets of an organization's operations, strategies, and culture, digital transformation aims to improve efficiency, customer experience, and competitiveness, while enabling organizations to adapt to rapidly evolving technological advancements and market conditions.

Though strategic imperatives evoke digital transformation initiatives, core constituents of any digital transformation program must identify its drivers and enablers (Table 1).

As can be inferred from the drivers of digital transformation, successful initiatives require a clear vision, strong leadership, adequate financial resources, and a willingness to change. It is also clear from the enablers of digital transformation that organizations must embrace change to choose, implement, and adopt effective technologies to be best positioned in a rapidly evolving digital landscape,

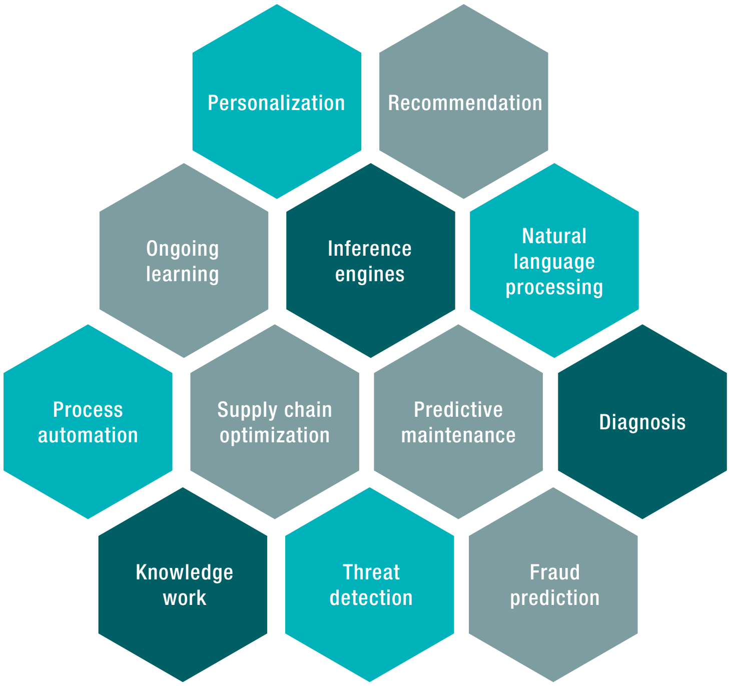
As digital transformation initiatives aim to improve competitiveness through achieving business process efficiencies in delivering superior customer experience, rapidly evolving technologies, such as AI, have a significant role as they have the potential to deliver sustainable competitive advantage to organizations.

## 3. WHAT IS THE ROLE OF AI IN DIGITAL TRANSFORMATION?

AI plays a crucial role in digital transformation by infusing the intelligence and automation to make the most of digital technologies. AI technologies mine and analyze different types of data in organizations and identify patterns to make inferences and recommendations. For instance, by analyzing customer data, such as products sold or services delivered, organizations can mine such data over a period of time to identify patterns in customer preferences.[3] These analytic

[3] Calp, M. H., 2020, "The role of artificial intelligence within the scope of digital transformation in enterprises," Advanced MIS and digital transformation for increased creativity and innovation in business, IGI Global, 122-146

**Figure 1:** Components of AI for digital transformation




insights create opportunities for personalized marketing campaigns, cross-selling related products and services, up-selling of frequently bought products and services, etc., AI technologies, such as pattern recognition algorithms, inference engines, recommendation engines, and predictive analytic engines, offer several ways in which AI contributes to digital transformation.[4] Components of AI that enable digital transformation (Figure 1) include:

- **Personalization:** by analyzing customer data, personalized customer experiences can be created based on consumption patterns, tailoring products, services, and recommendations to suit individual preferences. Personalization enhances customer engagement and builds customer loyalty.
- **Recommendation:** by gaining insights on user interactions in websites, e-commerce and digital platforms, recommendation engines can suggest relevant products, services, and/or content to users, increasing user engagement and sales leads.
- **Learning:** thorough repetitions, AI algorithms can continuously learn from data and adapt to changing conditions, ensuring that organizations remain responsive to rapidly evolving business environments.
- **Inference engine:** machine learning algorithms can process vast amounts of data quickly and extract valuable insights. These analytic insights enable organizations to make data-driven decisions, identify trends, and to understand customer behaviors with greater accuracy.
- **Natural language processing (NLP):** by analyzing language constructs, machines can process natural language to understand and interact with human language. Knowledge agents, such as ChatGPT, chatbots, and virtual customer services assistants, use NLP to help improve customer support and communication.
- **Process automation:** robots and AI can help streamline operations and handle routine tasks, reduce manual tasks and errors, and optimize business processes, freeing up employees for more strategic work.
- **Supply chain optimization:** by optimizing logistics and associated business processes, predicting demand, managing inventory, leading to cost savings and improved customer satisfaction, AI can optimize supply chain operations.
- **Predictive maintenance:** based on sensor data and historical information, AI algorithms can predict equipment failures or maintenance needs, reducing downtime and

[4] Davenport, T. H., and N. Mittal, 2023, “How companies can prepare for the coming “AI-first” world,” Strategy & Leadership 51:1, 26-30

improving efficiency, in industries like manufacturing and healthcare.

- **Diagnosis:** by analyzing patient data in healthcare, AI can assist in interpreting medical images, diagnosing diseases, and aiding in more accurate and timely healthcare decisions.
- **Automation of knowledge work:** AI knowledge agents, such as ChatGPT, perform tasks traditionally associated with human expertise, such as legal research, financial analysis, and even creative content generation.
- **Threat detection:** cybersecurity tools using AI algorithms can identify and respond to security threats in real time, helping protect an organization's digital assets and sensitive information.
- **Fraud detection:** by analyzing transaction data and identifying unusual patterns or behaviors, AI algorithms can detect fraudulent activities thereby enhancing security and reducing financial losses.

Overall, AI serves as a critical enabler of digital transformation, helping organizations harness the power of data, automate processes, enhance customer experiences, and sustain competitiveness. Capability of AI algorithms to learn and adapt continuously makes it an invaluable asset in the digital transformation efforts.

AI relies on digital neural networks, which are inspired by the structure and functioning of the human brain. They can mimic the way humans learn from data, making them a powerful tool for solving problems that involve learning and adaptation.

## 4. HOW ARE NEURAL NETWORKS SIGNIFICANT IN AI?

Neural networks are of significant importance in the field of AI. They can be adapted and trained for a wide range of tasks, from image and speech recognition to recommendation systems and game playing. Neural networks can be designed for continuous learning,[5] allowing them to adapt and improve over time as new data becomes available. This adaptability is crucial for AI systems that need to stay current and relevant. The versatility of neural networks makes them a foundational technology in AI (Figure 2). Neural networks can also scale to handle large and complex datasets. Modern deep learning models are designed to work with massive amounts of data, which are increasingly available in digital transformations in finance.

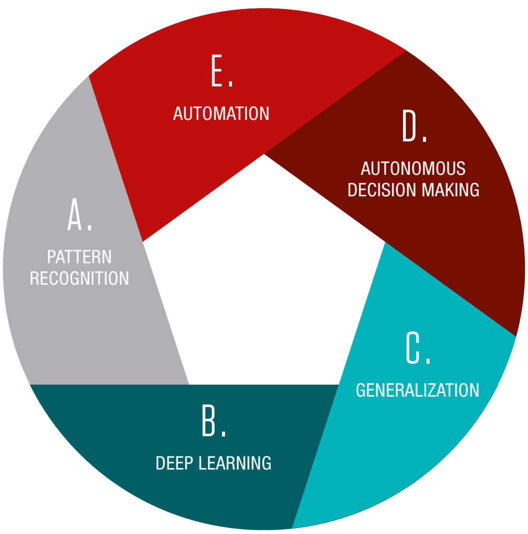
A. **Pattern recognition:** neural networks excel at recognizing and learning complex patterns in data, making them well-suited for tasks such as image recognition, speech recognition, and natural language processing. This pattern recognition capability is fundamental in many AI applications, including finance.

B. **Deep learning:** deep neural networks, also known as deep learning models, have revolutionized AI. These networks consist of multiple layers of interconnected neurons (hence the term "deep"). They can automatically learn hierarchical representations of data, which are essential for tasks like image and speech recognition. Deep learning has achieved remarkable success in various AI domains, from autonomous driving to natural language understanding.

C. **Generalization:** neural networks can generalize from the data they are trained on to make predictions or classifications on unseen data. This ability to generalize is a key characteristic of AI systems.

D. **Autonomous decision making:** neural networks can make autonomous decisions based on the patterns and information they have learned. This capability is essential for AI applications like autonomous vehicles, robotics, and natural language understanding.

**Figure 2:** Capabilities of neural networks




[5] Smith, J., 2021, "The significance of neural networks in artificial intelligence," Journal of Artificial Intelligence Research 10:3, 123-135

**Figure 3:** Origins of challenges in digital transformation


E. **Automation:** once trained, neural networks can automate tasks that would be labor-intensive or time-consuming for humans. This automation can lead to significant efficiency gains and cost savings in various industries.

Neural networks are a cornerstone of modern AI due to their ability to learn complex patterns (A, B), adapt to new information (C), and automate tasks (D, E) across a wide range of applications. Their effectiveness in handling large datasets and their capacity for continuous learning make them a driving force behind many AI advancements and innovations. Their applications are found in a wide range of disciplines, including finance (e.g., fraud detection and trading algorithms), healthcare (e.g., diagnosis and drug discovery), marketing (e.g., recommendation systems), and more. Their interdisciplinary nature makes them a valuable tool for solving diverse problems in varied digital transformation initiatives.

## 5. CHALLENGES IN DIGITAL TRANSFORMATION AND ADOPTION OF AI

Digital transformation and AI adoption bring about numerous benefits, but they also come with several challenges that organizations must navigate. Key challenges in digital transformation and AI implementation stem from data, people, systems, and organizational issues (Figure 3):

Data are ubiquitous in digital transformation programs that ensuring security and privacy is paramount. With the increased reliance on digital technologies and data collection, organizations face heightened concerns about data privacy and security. Protecting sensitive information from breaches and ensuring compliance with data protection regulations, such as GDPR in the E.U., is a significant challenge. In addition, ensuring quality of data and integrating is another challenge. Data used for AI and digital transformation initiatives must be of high quality, accurate, and consistent. Integrating data from various sources and formats can be complex and time-consuming, leading to challenges in data preparation. As such, establishing robust data governance practices, including cataloguing data, properly tracking origins, and controlling access, is essential for managing data effectively. Implementing these practices, ensuring data security and privacy, data quality and integration, and proper governance can be challenging.

People are the backbone of organizations, and as such ensuring skilled and talented professionals with expertise in AI, machine learning, data science, and cybersecurity can be a daunting proposition in digital transformation programs. In addition, transforming an organization's culture and processes can face resistance from employees who are accustomed to traditional ways of doing things. Effective change management strategies are essential to ensure that employees embrace new processes, technologies, and workflows. Moreover, AI systems can inadvertently perpetuate biases present in historical data, and as such ensuring fairness, transparency, and ethical use of AI is a growing challenge. Organizations must address skilling up, resourcing, change management, and bias and discrimination issues to ensure effectiveness of digital transformation programs, and maintain trust and avoid legal and reputational consequences.

Systems organizations use are not easily integrated with modern technologies. Many organizations have legacy systems that are costly and complex to replace or upgrade. In addition, implementing digital transformation initiatives often requires substantial investments in technology, training, and infrastructure that measuring and demonstrating a clear return on investment (RoI) can be a challenge, particularly in the short term. Further, as organizations grow, their digital infrastructure and AI capabilities must scale to accommodate increased data volumes and user demands. Ensuring scalability without sacrificing performance is another challenge. It is, therefore, essential for organizations to plan for the integration of legacy systems effectively, establish RoI targets, and ensure a scalable technology platform for the future.

Organizational readiness for digital transformation is a real challenge in terms of compliance with evolving regulatory frameworks, especially in industries like finance and healthcare. Staying abreast of regulatory changes and adapting systems accordingly is as essential as ensuring that concerns around the environmental impact of data centers and AI computations are managed in such a way as to find sustainable solutions to reduce their carbon footprint. In addition, managing risks such as reliance and/or dependence on a single vendor, specific technology, or platform, is a challenge. Organizations must, therefore, ensure a framework for regulatory compliance, sustainability, and risk management to be successful in their digital transformation initiatives.

Addressing these challenges requires a holistic approach that combines technology, leadership, organizational culture, and ongoing learning and adaptation. In the case of AI-driven tools and applications, user adoption can be challenging if users are not familiar with, or resistant to, AI technologies. Organizations that successfully navigate these challenges are better positioned to reap the benefits of digital transformation and AI innovation opportunities.

**Figure 4:** Opportunities with digital transformation


> *"Transforming organizations deliver superior customer experience and greater value to their shareholders by systematically studying and effecting changes to their people, processes, and systems."*

## 6. OPPORTUNITIES WITH DIGITAL TRANSFORMATION AND AI

Notwithstanding the myriad of challenges, digital transformation presents numerous core opportunities for organizations across various industries. As discussed earlier, these opportunities can lead to improved efficiency, customer experiences, competitiveness, and innovation (Figure 4).

Major opportunities associated with digital transformation come through data analytics, where organizations can gain **customer insights**, a deep understanding of customer behavior and preferences. As discussed earlier, these insights help refine marketing strategies, improve product design, and tailor offerings to specific customer segments.

Through data analysis and AI-driven customer insights, organizations can offer better product recommendations and improved customer support. As such, digital transformation enables organizations to provide personalized and seamless **customer experiences**. Digital channels such as social media, chatbots, and mobile apps enable organizations to engage with customers in real-time, gather feedback, and provide instant support, improving **customer engagement**. Moreover, digital tools can improve **employee productivity** and satisfaction by automating repetitive tasks, offering remote work options, and providing access to training and development resources. Digital tools also facilitate **collaboration** among employees, partners, and customers, enhancing connectivity, and communication tools enable remote work, global partnerships, and real-time collaboration, improving **employee productivity.**

**Operational efficiencies** are at the heart of any digital transformation initiative, as automation and digital tools streamline business processes, reducing manual tasks and the potential for errors. This leads to increased operational efficiency, cost savings, and faster response times, providing organizational **agility and flexibility** to respond to market conditions more easily. As discussed earlier, digital transformation can **optimize supply chain** management, reducing costs, improving inventory management, and enhancing overall efficiency. As stated above, predictive analytics and AI can help in risk identification and mitigation, helping organizations proactively **manage risks** and vulnerabilities. Moreover, digital technologies have been proven to reduce environmental impact,[6] contributing to **sustainability** goals, and appealing to eco-conscious consumers.

The greatest of all opportunities in digital transformation lies in **data-driven decision making**, as digital transformation initiatives allow organizations to collect and analyze vast amounts of data. As discussed, leveraging data analytics and business intelligence tools, decision makers can make informed choices based on real-time insights, improving resource allocations and strategic planning. In addition, organizations can explore opportunities to **monetize their data** by offering data-driven services or insights to other businesses.

Furthermore, digital technologies infuse **innovation** where companies can develop new products, services, and business models that cater to changing customer needs. This can result in the creation of **new revenue streams** and market opportunities. Consequently, organizations that embrace digital transformation gain a **competitive advantage**, adapting more quickly to market changes, responding faster to customer demands, and staying ahead of competitors who are slower to adopt digital technologies. Thus, digital transformation enables organizations to expand market reach and create opportunities for international growth, **reaching a global audience** for their products and services.

## 7. A FRAMEWORK FOR SUCCESSFUL DIGITAL TRANSFORMATION

Organizations transform themselves not only by investing in enabling digital technologies, but also by implementing the right strategies to ensure success in digital transformation. Visionary leadership, inspiring and leading change in every area of the business, is critical for realizing transformation objectives.

A useful LEAD-CHANGE-TRANSFORM framework (Figure 5), where strategic business leadership to ensure necessary changes are implemented in data, processes, systems, and people is proposed to enhance the degree of success in digital transformation initiatives.

**Figure 5:** A framework for successful digital transformation

<table>
<tr><td>LEAD</td><td colspan="4">Strategic business leadership<br>Strong goal-oriented leadership on digital transformation initiatives from the senior management</td></tr>
<tr><td>CHANGE</td><td>People<br>(Talent acquisition and employee engagement)</td><td>Processes<br>(Optimized value chain and end-to-end process automation)</td><td>Systems<br>(Proprietary and off-the-shelf tools and technologies to best-fit solutions)</td><td>Data<br>(Mission critical data for information portal)</td></tr>
<tr><td>TRANSFORM</td><td colspan="4">Digital transformation<br>• Sector-leading<br>• Profit margin/value creation<br>• Greater growth and earnings<br>• Improved shareholder returns</td></tr>
</table>

[6] https://tinyurl.com/yckbwwzv

## 8. CONCLUSION

Digital transformation offers a wide range of opportunities for organizations to improve their operations, stay competitive, and deliver enhanced value to customers. These opportunities span across multiple aspects of business, including customer engagement, operational efficiency, innovation, and global expansion. Digital transformation and AI adoption come with several challenges that organizations must navigate. Despite these challenges, digital transformation presents numerous core opportunities for organizations across various industries to gain a sustainable competitive advantage.

# IS ACCOUNTING KEEPING PACE WITH DIGITALIZATION?

**ALNOOR BHIMANI** | Professor of Management Accounting and Director of the South Asia Centre, London School of Economics[1]

## ABSTRACT

Digital transformations are taking place across enterprises in every industry. Becoming digital is both essential to compete and virtually unstoppable. All previous major technological disruptions have led to financial intelligence being altered to ensure more effective decision making in the face of change. This article considers issues that organizations going digital need to address in relation to accounting information provision. It discusses several points: accounting's need to move toward the delivery of predictive information rather than relying on extrapolations of historical data; the recognition that machines make more decisions that alter accounting information needs, structures, and contents; the importance of recognizing the "data-learning-action" loop that is emerging; the emergence of "strat-perational" information contexts; and the relevance of prioritizing qualitative insights in decision making.

## 1. INTRODUCTION

Three decades ago, new computer technologies changed how businesses manufactured products and delivered services. Accountants, as a result, altered the information they provided to managers to enable decisions to be made that were aligned with the new business environment. Digitalization today is not only transforming products and services but also the form and type of information enterprises must grapple with. Consequently, the interface evolving between digital technologies and financial information is likely to become the biggest shift ever seen in the history of business decision making. The beauty of digital is that it unfolds novel products, services, and value creation possibilities while also changing the nature of information managers can use to steer their organizations. The question remains: are enterprises deploying financial intelligence that is up to the mark? This article examines key financial information changes that digitalization necessitates to ensure decision makers are not retaining a mindset belonging to conventional analogue modes of operating.

## 2. THE LIMITS OF ACCOUNTING

Accountants take the view that there exists no business situation that accounting cannot report on. From cost determinations, to auditing, to taxation, to financial analyses, accounting information is considered to provide a sound basis for assessing financial performance and to make business decisions. However, accounting today is facing an unmitigated rupture – it needs to restructure itself from the core. While financial information within balance sheets and income statements will likely always matter at some level, accounting's sole focus on past economic transactions and business outcomes cannot remain. Several disruptive forces confront modern accounting expertise. To start, accounting information needs to address what will take place rather than simply reporting on what has transpired. The finance professional must veer toward delivering predictive insights based on a wider analysis of data as opposed to focusing on historical accounting reports perused by decision makers who must then extrapolate the implications. Going digital provides executives a bridge to tomorrow as data becomes more

[1] The ideas in this article are drawn from my books: Bhimani, A., 2021, Accounting disrupted: how digitalization is changing finance, AICPA/Wiley, and Bhimani, A., 2022, Financial management for technology start-ups, Kogan Page

predictive of financial outcomes to follow rather than ones that have already materialized. To achieve this, accountants must develop skill sets enabling them to assess more diverse organizational datasets.

Secondly, many firms are seeing a greater portion of executive action being taken through autopilots without human input. As such, accounting reports that used to guide such actions are becoming redundant because machines do not need to "read" reports formatted in a specific way. Machines act and operate simultaneously. In effect, the forms and intents of accounting information are changing because the agents requiring financial reports are changing. The notion that humans must make decisions based on information and only then act is inimical to the manner in which machines function, since information retrieval, analysis, decision making, and action are not dissociated processes.

A third key trend is that digital technologies increasingly enter products as ID tags and IoT devices, whereby information systems and data gathering mechanisms are baked into the products themselves that are being reported upon. Accountants have traditionally collected data about processes and products using information systems developed to capture data and convert these into useful information. They now need to focus on unraveling new insights from products that are themselves also information systems. Furthermore, it is not just a matter of products and information systems being intertwined but contexts where digital technologies structured on blockchain applications that record transactions in a manner that grounds the assurance of transactional legitimacy. In such organizational spaces, trust becomes integrated as part of information content, placing the traditional role of audits under question.

Accounting is without doubt becoming a whole lot more complex and digitalization is at the heart of the ongoing disruption. In altering what information sound decision making must rest on, digital technologies are crowding out conventional business philosophies, models, and thinking. Their capacity to self-transform further displaces the traditional role of accounting, premised on a linear sequence of data collection followed by conversion into financial information that accountants produce in a digestible form, allowing humans to base their actions upon. Business history knows no such pace or scale of change and there exists no U-turn to this technological transformation. Executives cannot afford their accounting information to stand still, as digitalization progresses without relent.

> "*...the interface evolving between digital technologies and financial information is likely to become the biggest shift ever seen in the history of business decision making.*"

What is of greatest relevance to business today is acknowledging the ramifications of going digital. It has been said that we have entered a new industrial revolution. Two hundred and fifty years have passed since mechanization emerged in industrial activities. A hundred years ago, we saw electrification and mass production taking shape, and electronics and automation started a third revolution around sixty years ago. The ongoing "fourth" industrial revolution has ushered in a convergence of our biological, physical, and virtual worlds. But, of note, is that during the first three revolutions, few people were cognizant of the magnitude of changes to come. Today, however, we are all too aware of the massive and unprecedented shifts that are ongoing in the way we produce, consume, move, communicate, and experience things. While the industrial changes surfacing during the first three industrial revolutions could perhaps not be foreseen by people, no one today can stand in denial of the all too evident and fast-paced changes this fourth one has unleashed. No excuse can be had for financial information to not react, or indeed proact, in this climate of extreme change for business.

## 3. UNKNOWNS IN CORPORATE STRATEGY EXECUTION

Executives face an increasing array of unknowns. In the past, decision-makers pursuing major alterations in their operations fully appreciated what they were seeking. It could have been a novel marketing campaign to rebrand a product, implementing flexible work practices to make production more pliable, possibly facilitating customer service via new support systems, perhaps making capital investments to increase productivity and scope, or, possibly, to seek a merger for acquiring new knowledge and mobilize wider revenue streams. Decision-makers have had the luxury of reasoning and purposefulness in investments they made in pursuit of specific business outcomes. Going digital, by contrast, implies operating within a limited vision of what the end state of executive action might be. There are no mechanisms available that can lend assurances that a specific digital transformation investment will culminate in specific enduring outcomes. Indeed, that is the point of digital: pursuits leading to outcomes that are more dispersed than the ones that were initially advanced, and which may even be unintended, can lead to strategic moves that are not only viable but extremely apt and which trigger the next move. To a degree, digitalization enables digitalization that sets off new trajectories, which alter processes, all the while enabling further action that is far from being anticipated at the outset. Speed of action has always been of relevance in business, but going digital powers fast iterative changes that executives may not fully fathom and, which, therefore, remain outside what competitors can envision. Digital paths of action can take enterprises to different destinations, where reaction and proaction are essential to effectively and continually address the status quo, whose half-life becomes more transient.

We turn now to how accounting must be re-thought.

## 4. WHAT NEEDS TO CHANGE

How data can assist managers is changing. Financial information systems capture data points about economic transactions that have occurred, converting them into information that is condensed, formatted, and made intelligible to decision-makers. Digital data is much broader than economic data points. We can think of conventional internal accounting reports as being intended for executives through the collection of economic transactions-based data and related quantitative and qualitative metrics to feed into their decision making (Figure 1).

Accountants within enterprises that are digitally transforming must have greater engagement in assessing cloud infrastructure benefits, as well as mobilizing process changes and containing their costs, assessing cybersecurity constraints, and exploring pathways to more flexibility, automation, and scalability. Their role is becoming increasingly complex in relation to dealing with more varied data sources and data volume and becoming more cross-functional with decisions having to be made on enhanced rules-based automation across organizational activities.

**Figure 1:** Accounting drives decisions


**Figure 2:** Structured and unstructured data counts

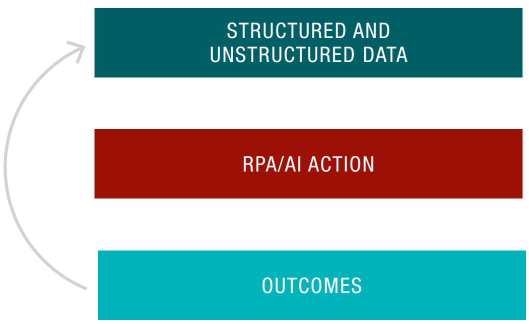
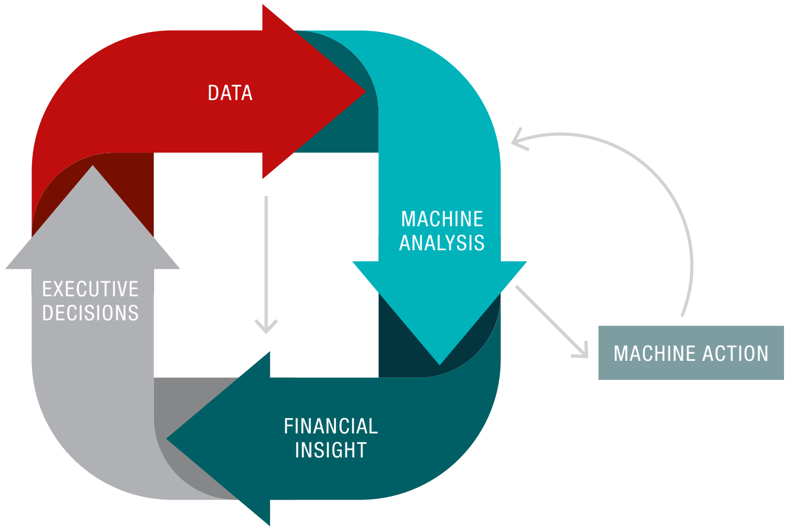


**Figure 3:** Executive action by humans and machines




Digital transformation offers a much wider realm of insights for executives to benefit from and accountants need to become adept in determining what maximizes the insights they can advance for the enterprise. Data that is external to the domain of economic transactions can enable the identification of trends that foretell economic exchanges and provide intelligence on novel possibilities for action and opportunities for growth. Executive action based on such data creates more such data that can be acted upon. Technologies such as robotic process automation (RPAs) are becoming increasingly sophisticated, with machines undertaking devolved action and continuously learning from those actions and refining information, yielding insights for assessment that executives can in turn harness (Figure 2). Financial information systems need to react to the possibilities that non-financial data has implications for consequent financial transactions. Beyond that, machine-based actions culminate in decisions that executives can undertake based on the growth of structured, financial, and non-financial, as well as unstructured, data (Figure 3).

With machine learning, machines can assist in taking and operationalizing certain actions reliant on data inputs and subsequently to refine these actions based on data outputs from prior actions. Concurrently, humans can focus on using information to guide actions they do not delegate to machines. Actions drawing on data insights from machine executions can become further integrated with human action related data outputs, producing a breadth of data that circularly underpins yet more decisions and actions. Ultimately, data growth enables decision making that fosters more data, enabling greater and faster learning, which in turn unleashes enterprise action that propels yet more data (Figure 4). That looping of data-learning-action is among the most powerful impact that going digital can trigger, if effectively channeled into intelligent decisions and action, whether by humans or machines. In contexts where artificial intelligence (AI) agents find increased presence, data becomes a principal basis for ML systematized learning. Consequent action based on that learning unleashes more data faster, enforcing even greater and faster learning. This gives form to a virtuous loop with extreme data analysis, accelerating learning, and extensive action powering the looping. It is this in AI intense environments that enables extremes of growth, potentially leaving competitors in rear view. As firms implement AI systems, the interfacing of data, learning, and action makes accounting more complex than it has ever been, while also enabling it to become more strategically relevant for organizational growth and performance.

Conventionally, strategy is regarded by some executives as emerging over time, as intentions collide within their firms. But for most, strategy tends to be formulated to define and guide desired operations, taking account of wider business and environmental factors. As such, strategic information points executives to what should underpin their decisions in pursuit of targeted business activities, with such decisions having to take account of marketing advances to be made, investment opportunities to be identified, branding efforts to be made, and so on. And, of course, operations should essentially tally with, and support, strategic intentions. Within digitally transforming organizations, it may well be that the past should not determine strategy.

Just as this article argues for accounting information to be predictive and focused on the present to point to what is to follow financially, strategic action can no longer be defined based on what has happened. But a further point must be made, which is that it is now inappropriate to see strategic intent as presuming a long-term time period, so the enterprise paves its protracted directional purpose. In rethinking alternative courses of action on a continuous basis, executives cannot adhere to a pre-digital conception of strategy. In fact, strategy and operations are intertwined in digital contexts, so much so that strategic action that is now of essence is one that recognizes the process of data capture, decision making, and action having become "strat-perational" (Figure 5). Strategy has moved to the "here and now", directing organizations to consider whether operations need refining and also whether strategic intents need redefining in an interrelated realm. This implies that accountants should no longer lock-in expired rationales into digital contexts, because the information potential of operations intermeshed with strategic aims goes much beyond treating the two as separates, with long-term strategic intent guiding day-to-day operations.

Management thinkers have long held that enterprise controls should keep the strategic apart from the operational. The idea of short-term actions being informationally divorced from long-term aims is one to be questioned when going digital. In digital, the short-term and the long-term enjoy a close coupling that has not been present in conventional industrial businesses. Within digitally transforming contexts, a decision to pivot the firm in a certain direction will create data to be analyzed in real time. Aggregate data produce information that enters decisions concerning the next move or desired action. Under digital, broad business pursuits meet with continuous realignment. Operations can produce sufficient information to suggest the enterprise should steer toward entirely novel areas of activity or a new direction. The broad business hypothesis should be open to being questioned and altered in the face of information from operational activities. Small sets of data cannot be maintained as islands distanced from pointing out alternate organizational agendas. Managing in real time and quantifying short-term returns must naturally preserve relevance, while a recognition must be had also that longer term changes are interdependent, with ground level operational processes and business direction being flexible to pointers that are operational. What some view as an "agile business" in digitalizing enterprises, is in fact a rejection of demarcating operational activities as entirely separate from strategic moves. Digital transformation means the two are coupled with differentiation being an obstacle to sound management. This is not to suggest that thinking strategically

**Figure 4:** Data enables learning, which guides action

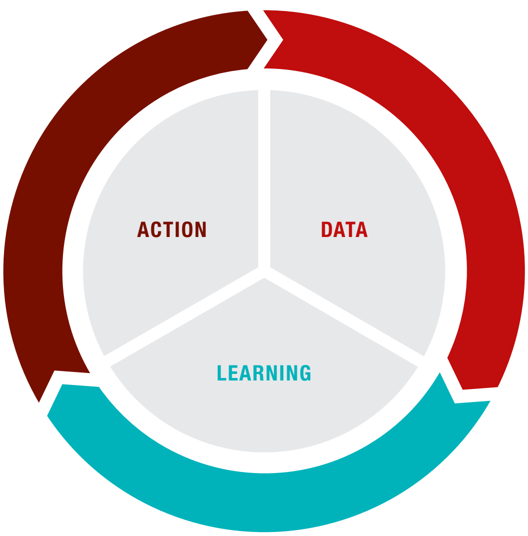
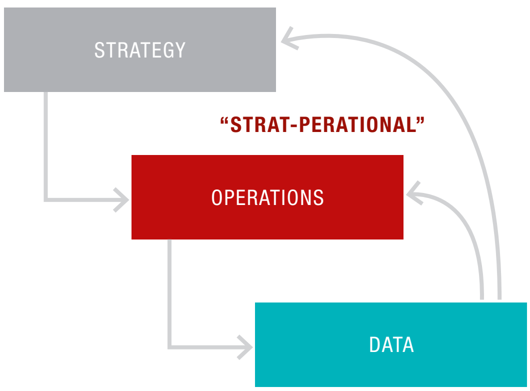



**Figure 5:** When strategy is truly coupled with operations



no longer has a place, but that distancing such thinking from the value of real time action and operational on-goings is to deny the production of insights that can be important to survival and growth.

## 5. CONCLUSION

Digitalization is bringing a multitude of changes and challenges that are impacting what executives need from their corporate finance function. The accountant must comprehend the significance of shifts evolving at the interface of digitalization, financial information, and enterprise action. Clearly, as organizations invest in digital technologies, hordes of new data types emerge, not all of which is financial in a conventional sense. This article has discussed several implications including, first, that accounting information must now steer toward being predictive rather than forcing decision-makers to try to visualize the future consequences of their actions based on reports of what has transpired. Second, it has been noted that machines are increasingly engaged in integrating data analyses with decisions and actions, such as to alter the information roles, content, and needs humans may have placed on accounting. Thirdly, it has been discussed that being cognizant of the "data-learning-action" loop is of high relevance, as this is what going digital rests upon. It has also been argued that differentiating between long-term strategy and short-term operational activities can no longer be seen in clear-cut terms; "strat-perational" information contexts are fast emerging, which impact what accounting information should focus upon. In addition to these observations, one more point needs to be borne in mind, which is that the relevance of quantification in business decision making must not be overstated simply because digitalization is taking place. Data-driven management action is certainly desirable, though the implication cannot be that numbers should trump qualitative assessments. Going digital cannot pre-suppose that numbers-based analysis should monopolize enterprise decision making. In the face of information growth, senior executives rely on more, rather than less, qualitative input when businesses digitalize. The movement towards digitalization certainly increases the possibilities for numerical analyses, but this signifies more, rather than less, qualitative insight. Understanding what going digital means for their enterprise should be a priority for organizational leaders and this will involve unlearning at least some conventional managerial precepts that have been said to be fundamental, including the idea that managing by the numbers should be prioritized. Financially relevant information should not preclude that which is most capable of generating insights, and, as it turns out, much of that cannot be quantified.

# BANK AND FINTECH FOR TRANSFORMATION OF FINANCIAL SERVICES: WHAT TO KEEP AND WHAT IS CHANGING IN THE INDUSTRY

**ANNA OMARINI** | Tenured Researcher, Department of Finance, Bocconi University


## ABSTRACT

Technology in banking has always had the power to affect the fundamentals of business, such as information and risk analysis, distribution, monitoring, and processing. The relationship between technology and banking is, however, quite different to how it used to be, predominantly due to stronger interdependencies, both technological as well as strategic. Today's digital technologies have the power to improve efficiency and effectiveness in services, as well as exerting increasing influence on banks' products and delivery methods, and increasingly on strategies. Digitalization is changing the rules of the game in many industries, and this results in the emergence of complex and dynamic ecosystems for growth and innovation. The main forces shaping these changes have led the financial services industry to reconsider the role of banking and finance, more as an "enabler" for many other businesses and commercial initiatives than as a mere provider of products and services. This paper looks at how financial services organizations are transforming themselves using the new technologies at their disposal and tries to determine what should be kept and what needs to change.


## 1. INTRODUCTION

In recent years, the term financial technology, or fintech, has emerged as a key driver for most of the changes in the financial services industry, even though technology has always been an enabler. This is because the core business is made of services and services can be produced and delivered to processes with a high degree of technological intensity.

Technology in banking has always had the power to affect the fundamentals of business, such as information and risk analysis, distribution, monitoring, and processing [Llewellyn (1999, 2003)]. However, it is useful to make a distinction between technologies of the past and the digital technologies of the present. The latter not only have the power to improve efficiency and effectiveness in services but have also started to exert increasing influence on banks' products and delivery methods [ECB (1999)].

Consequently, if we think of fintech there are two meanings worth considering. Firstly, there is the implementation of technology in the industry to improve efficiencies in the back end of a single financial institution (payments infrastructure, for clearing and settlement, as well as offer new services to customers, such as ATMs), as well that of stock market infrastructures to increase buying and selling securities in general. This is the "old" fintech. Secondly, because of the nature of immaterial components of financial services, which makes it quite simple to distribute globally as well as to develop many innovations by simply unbundling and re-bundling solutions, they are the most successful evolutionary entities and disruptors. They combine multiple sources of value – such as cost efficiency and customer experience and engagement – to create disruptive new business models and exponential gains. This is the second definition of fintech. FSB (2017) describes fintech "as technologically enabled innovation in financial services that could result in new business models, applications, processes, or products with an associated material effect on financial markets and institutions and the provision of financial services. Fintech innovations are affecting many different areas of financial services."

However, both have paved the way for technological transformation, which over time has become a key priority for many banks and financial services institutions that were seeking to remain competitive and meet customers' evolving demands. More recently, tech changes have driven the market into a digital vortex, where organizations are forced to compete in a digital environment in which business models, offerings, and value chains are digitalized to the maximum extent possible. This has led to the creation of new disruptions and blurring of the lines between industries. The advent of new ways of doing business, such as "platform-as-a-service" (PaaS), "software-as-a-service" (SaaS), and more specific to our case, "banking-as-a-service" (BaaS), are helping to lower the barriers to sophisticated financial applications by allowing people and talent to focus on business value added tasks as opposed to the building, supporting, and managing of infrastructures.

Digitalization is changing the rules of the game in many industries, and this results in the emergence of complex and dynamic ecosystems for growth and innovation.

The main forces shaping these changes have led the financial services industry to reconsider the role of banking and finance, to become more of an "enabler" for many other businesses and commercial initiatives (such as the embedded finance phenomenon) than a mere provider of products and services. And there is a growing market awareness of the role that technology is playing with regards to this vision.

At this stage, there are two key issues that are crucial in the quest for digital transformation of banking and financial services. The first pertains to key trending topics that shape the industry, while the second revolves around the solutions and platforms employed to enable these trends. With regards to the trending topics, there are a few of them, such as sustainability, segment-specific banking, digital financial advisory, digital identity, instant payment cryptocurrencies, central bank digital currency (CBDC), and open banking and finance, etc.

Given the rapid pace of change, what industry leaders need to consider is: where does the new come from, and what should we keep from the old?

The new comes from the way the market and its actors are looking at customers. For many years, banks have driven their activities from the perspective of segment-specific banking, where banking interfaces – such as branches, websites, e-banking portals, and mobile apps, more recently – have typically been designed with a single interface intended to cater to multiple customer groups.

This approach fails to consider the diverse needs, preferences, attitudes, and behaviors of customers. In specific, it fails to consider the changes in behaviors, expectations, and attitudes that accrue as a result of socio-economic changes and supply and demand cycles of innovation. Banks also ignored the fact that functional requirements are only one of the reasons why customers make a choice, and that different customers have different needs and expectations. Even simple issues, such as the differing expectations between different age or wealth groups, are typically ignored. Addressing the unique needs of various customer segments, such as private banking or high net worth clients, will need to involve some form of digital transformation.

The second key dimension of digital transformation relates to the solutions and platforms employed to enable these trends. Value chains need to change from pipeline business models to platforms business models where unbundling them into different modules of products or services can develop new value propositions.

The financial services industry is facing new waves of change because of these technologies, and banks are the most affected by many of these changes. They are facing competition not only from their conventional peers, but also, due to their customers becoming more knowledgeable and demanding, with fintechs, bigtechs, and even retailers. This situation is exacerbated by the fact that customers are constantly changing their attitudes, behaviors, and habits.

We are currently in the early stages of transforming the banking sector and the implementation of new technologies, and both regulators and supervisors must also face the additional challenge of digital transformation, which requires achieving the right balance between promoting new digital value propositions and protecting customers and banks against the risks inherent in digitalization of financial services. Under these circumstances, there are old and new risks, but also old and new managerial principles and rules to detect and become aware of.

The new enlarged non-financial risks (for example, fraud, cyber risk, operational and strategic risks, etc.) come from the increasing use of cloud, big data, platforms, artificial intelligence (AI), machine learning, and other seamless tools, which aim to increase personalization and improve user experiences to deepen relationships. While old risks remain, they may even become exacerbated by the new complexities.

If we then consider managerial rules, we need to recognize that there are core principles – such as the basic principles of keeping safe economic, financial, and patrimonial balances – and that simply because business is undertaken it does not mean the business has changed.

The new is the digital layers that lie on top of the old ones, making value chains looking – only apparently – shortened. However, they introduce a kind of infinitive intermediation in the market [Omarini (2019)] when the old infrastructures are still in use (think of some services of payment).

New banking is moving its business model framework from vertical silos to platforms and richer ecosystems. The result will be the de-integration of financial solutions, which can easily become embedded into the business value chains of others. This new trend comes from the increasing adoption of modularity, which drives companies to move towards product componentization [Tuunanen et al. (2012), Accenture (2021)]. All of this is not completely new to the financial services industry, because banks have always developed bundles of services. What is new is that modularity and componentization require a deeper understanding of the customer base that is available today to design and develop customization and personalization for selected needs and use cases for the market [de Blok et al. (2014), Moon et al. (2010), Bask et al. (2011), Silvestro and Lustrato (2015), Bleier et al. (2018), Anshari et al. (2019)].

The new also comes from the digitalization of the many processes that have been re-designed, as well as the need to make a number of older products match the tech-driven pace of change.

The new current outlook reveals nascent ecosystems made of independent actors, where the traditional supply-centered oligopoly is coupled with fintechs, techfins, retailers, etc. Within this also lies the disruptive aspect of PSD2 (Payment Service Directive 2) and 3 for Europe, and similar initiatives in other major markets.

PSD2, which seems to be going one step beyond its regulatory mandate [Cortet et al. (2016)], is indeed an impressive accelerator of the digitalization process that is already affecting banking. In fact, it aims to boost competition in the market, and, therefore, customer mobility, by increasing unbundling and modularization in the industry. It is also challenging the financial services landscape, and its stability, by severely impacting revenue streams that were considered sticky by banks.

The difference with the past, in terms of the relationship between technology and banking, is the stronger interdependencies: technological as well as strategic interdependence.

## 2. THE DIGITAL TRANSFORMATION OF THE CONSUMER

When thinking of the consumer, we must bear in mind that what they demand and expect from banks can only be partially defined in financial terms. Indeed, they want their life to be easy and the path to their goals to be a simple one. They look for convenience, product simplicity, ease of use, cost savings, personalized offerings, and memorable experiences [Omarini (2019)].

The COVID-19 pandemic has further incentivized customers to shift away from traditional branches of banking towards digital channels, placing the industry at a turning point. For a long time, the main objective of most banks was to increase their share of the customer's so-called "share of wallet", which meant that banks tried to get as large a share of a customer's wealth as possible. However, over time, some large institutions have shifted their activity from deposit-taking, lending, securities underwriting, and trust services toward dealer and market-making activities, brokerage services, and proprietary trading. The result has been a fall in customer-centricity; hence traditional banking has lost significance vis-à-vis other forms of financial intermediation and counterparts (namely fintechs, bigtechs, etc.).

While banks have traditionally been convenient one-stop shops for businesses and consumers, they are currently undertaking this digital shift differently. On the one hand, there are several banks that have not evolved their products in a way that matches the tech-driven pace of change in other industries. On the other hand, there are others that still find it difficult to undertake this change. A recent ECB study (2023) found that banks – under its supervision – still find it difficult to isolate and quantify the cost and revenue impacts of their digital transformation strategies and processes. The study also found that half of the sampled banks do not monitor the number of customers digitally onboarded; only one in four banks can quantify the volume of digital sales. Only half of the banks monitor the number of digitally concluded loans (e.g., pre-decided loans, consumer credit), which stands at around 45% of their total loan portfolio.

This means that there is still a lack of vision regarding how digitalization is impacting the competitive landscape and which organizations are the game changers that are playing different roles in the market arena, so that they may be both competitors and partners.

Of course, it is not all bad news. There are a number of banks, such as the Royal Bank of Canada, ING, BBVA, that have embraced digital transformation, have declared renewed and different visions, and driven their business models to change.

Products such as checking accounts, loans, and even corporate advisory can only seem to be undifferentiated. If this is the case, then people may increasingly feel frustrated by the financial fragmentation that banks have imposed on many consumer processes and related experiences. For instance, buying a home once required navigating a complex environment of disconnected real-estate brokers, mortgage lenders, insurance companies, lawyers, and so on. At that time, the bank-customer relationship was driven by the branch's manager, and the face-to-face relationship helped mitigate a number of concerns/issues. Furthermore, there were not that many alternatives in the market with whom the borrower had a long-term banking relationship with. Today, there are many new ways to reach and connect with consumers, and banks need to identify and engage with these customers, as their newer competitors are doing.

Everyday banking encompasses day-to-day financial services, such as checking and savings accounts, credit cards, personal loans, payment processing, and lines of credit on the traditional-banking side, for individuals and for small and medium-size enterprises (SMEs). This arena will also include e-commerce ecosystems, loyalty programs, discounts, advertising, and peer-to-peer marketplaces; meaning that banking need no longer be an obligation but something easy and even potentially enjoyable. Think of the fully-fledged e-commerce bank: Kaspi from Kazakhistan. Kaspi's customers have access to millions of products from more than 400,000 partnering merchants, ranging from low-price clothing and cosmetics to higher-price electronics, furniture, and jewelry. It enables easy, discounted shopping at retailers. Kaspi charges its partners a 5 to 11% fee, and its users pay nothing. For frequent purchases, they get cash bonuses deposited directly into their Kaspi accounts; a strong incentive to make Kaspi their primary bank [McKinsey (2022)].

> *"The goal of everyday banking is to be contextual and invisible. However, keep an eye on the other side of the coin: transparency and consumer protection."*

The common thread running through all day-to-day financial services is that customers want them to be hassle free, reliable, highly automated, and inexpensive. The goal of everyday banking is to be contextual and invisible, which means offering services that are cheap, easy, and accessible through many channels, such as the use case of "buy now pay later" (BNPL).

In combination, the above factors are fundamentally transforming the industry, resulting in increased competition, and as a result, falling profit margins for banks [KPMG (2016, 2023)]. If you consider that the average global banking return-on-equity (RoE) has fallen from 15% in 2008 to around 9.5 percent in 2021 and could potentially fall to 7.2% by 2030 [McKinsey (2022)] it quickly becomes clear why banking executives are so concerned.

Banking executives can no longer focus solely on costs, product and process quality, or speed and efficiency. They must also strive for new sources of innovation, creativity, and revenues. These increasingly complex forms of competition have forced banks to find new ways to attract and retain customers, who now, more than ever, command the power to choose [Omarini (2015)]. This new paradigm presents a formidable and constant set of challenges.

Image generated by Adobe Firefly

Customers are increasingly informed of what options are available to them, which in turn has led them to become more demanding. Thus, the paradox of the twenty-first century economy is that on one hand, consumers have more choices, which might yield less satisfaction, and on the other, top management has more strategic options that yield less value [Prahalad and Ramaswamy (2004)]. Hence, knowledge of what customers want is increasingly driving this new emerging paradigm.

The fact that digital technologies are changing the habits of how individuals do their banking will concern both banks and new players. They need to decide whether to take a purely transaction-driven business approach, which will allow them to survive under certain circumstances (such as volume, economies of scale, etc.), or a more relational-driven business approach, which will mean continuous innovation, boosted and driven by new ways of data management.

The current outlook for the banking industry reveals a network of platforms and a set of nascent ecosystems approaches [Breidbach et al. (2014)] made of independent actors, where the traditional, supply-centered oligopolies are coupled with fintechs, bigtechs, retailers, etc.

Within the new open banking framework also lies the disruptive aspect of PSD2 in Europe, and similar trends in other jurisdictions. Open banking provides "access to account" and communications with authorized third parties, customers, and payment account information. BIS (2019) defined it as "The sharing and leveraging of customer-permissioned data by banks with third party developers and firms to build applications and services, including for example those that provide real-time payments, greater financial transparency options for account holders, marketing, and cross-selling opportunities."

This is only the starting point, and there is no shortage of ideas regarding the challenges that banks face and the strategies they need to undertake in order to respond [Accenture (2018a, 2018b, 2020), AT Kerney (2021), CapGemini (2019, 2020, 2021), PwC (2018a, 2018b), Deloitte (2017, 2020, 2020a, 2020b), EY (2017), KPMG (2020), McKinsey (2017), Microsoft et al. (2017), Zachariadis and Oczan (2016), Dratva (2020)].

Banks have a number of options. They can view open banking and open finance frameworks (which, from June 2023, also allows sharing of information regarding mortgages, insurances, etc.) merely from a compliance perspective, or think of them as new competitive frameworks to develop. They can also expand their business lines (think of BaaS) or even transform their strategies and related business models [Cortet et al. (2016), Omarini (2022, 2023)].

Open banking and open finance are allowing new players to thrive not only in the payments area, but also in other areas of banking as well, once they have access to account and not-account information. This disruption is key to the ever-increasing unbundling and modularization of banking. While all the necessary conditions are already in place for the re-bundling stage, where the core objectives of financial intermediation may remain the same, the methods and functionaries relating to those objectives change with digital technologies and market developments. Think of the banking-as-a-service (BaaS) paradigm, which is driving endless possibilities, paving the way towards a truly embedded finance environment.

In this regard, BaaS unlocks new values because it allows banking to be embedded in adjacent ecosystems. It is the opportunity to eliminate the frictions in user interactions, among clients or partners, that make financial services more and more contextualized.

A high degree of open innovation [Chesbrough (2011)] is the result of the above. And the way banking has started being embedded in many other business value chains has also started empowering consumers to access not only their accounts, but also their mortgages, credits, student loans, automotive finance, insurances, investments, or pensions and loans. Ultimately, this access allows for the delivery of additional value in the form of saving-related services, identity services, more accurate creditworthiness assessments, financial inclusion, and a more tailored financial advice support service.

Opportunities associated with BaaS are taking the retail banking sector by storm, as organizations search for not only new ways of improving customer engagement and enhancing experiences, but also finding new sources of revenues from within and without the financial services marketplace [Finastra (2022)].

The era when all financial services were dominated by monolithic banking entities is over. We believe that the future of banking will be contested by banks and nonbanks in different arenas. An example would be everyday banking where payments, small savings, and consumer loans are core services, but that investment advisory, complex financing, and BaaS are also available.

Moreover, considering that in such a changing environment retail banking is increasingly in the business of being chosen [Omarini (2015)], being customer-centric requires shifting from a product-oriented view of business to a more service-oriented one, as the latter requires focusing on how the customers make "use of banks and banking" and not on the characteristics of the products. In this regard, customer intimacy can be a potential future direction because it aims to continually tailor, shape and re-shape products and services to fit an increasingly fine definition of the customer's expectations and needs.

Under the new customer intimate and digitalized approach, it is mandatory to look for new segmentation bases and criteria, because the success of both open banking and open finance will depend on customers being prepared and educated in becoming engaged, and willing to allow third-party providers to have access to their financial data. And therein lies that old core principle of "trust", which had been driven the financial services industry since its beginning and which will continue to matter to customers.

## 3. BANKING AND FINANCIAL SERVICES PLATFORMIZATION

Times have changed, and not even one of banking's main products has remained exclusively in the hands of banks or other conventional financial intermediaries. The banking business is one that is undergoing major transformation, as many of the boundaries between it and potential competitors have collapsed. New players in the banking industry have different understandings of what customers value and are more committed to customers than traditional financial service providers.

This huge change is being driven by new potential functionality, which is also spreading at the societal level [Alijani and Wintjes (2017)], where the borderless extension of financial innovation is experiencing great change and where the new fintech phenomenon has started developing and reshaping the industry's value propositions and related business models [IMF and World Bank (2019)].

All this will, in turn, accelerate the fragmentation of the value chains in the banking sector; as mentioned before, consumers are free to choose services provided by a set of third-party providers on the basis constituted by the (open) account they hold within a bank. This shift requires that everyone becomes aware of the need to move their mindset and related strategy from controlling to managing customers' money [Bareisis (2013), Omarini (2019)].

This means that the focus for every organization must shift from the value chain and the company's value proposition to the different ways value for customers can be developed and enriched over time. In this regard, vertical or pipeline business models may not be that good at satisfying the increasing customers' expectations, because they are all becoming very good at "comparing and contrasting" different offerings. Hence, organizations must look for a more holistic approach to customer knowledge and customer value. Providing customers with solutions through digital platforms is also transforming the banking business into business platform-based ecosystems, within which entities create value for one another by producing or consuming goods and services that mutually support one another.

In the first stage, banking is moving onto digital platforms; cross-industry interconnections will increase and result in new competitive threats. Providers of banking services will progressively come to see themselves in the role of "enablers" of transactions occurring on digital platforms and within business ecosystems.

For retail banking, especially in Europe, the advent of digital platforms can be expected to cause a shift away from the traditional universal banking business model towards a re-new customer-centered universal banking model. In the former, economies of scale and scope dominated strategic thinking, and conflicting of interests between business sections arose easily within the same legal entity. In the latter, the unbundling and re-bundling of services and respective business models are first selected and then chosen for a given purpose (such as solving a customer's need, improving quality, developing a new customer experience, etc.).

At this point, it is important not to confuse "platform" and "ecosystem". Platforms create value by eliminating frictions from transactions and exchanges; for example, in the case of a marketplace. The concept of ecosystems has become increasingly popular in several streams of literature (e.g., strategy, organization, innovation, digital models). The notion was first pioneered by Moore (1993), who referred to ecosystems as cross-industry entities. According to Moore's characterization, companies both collaborate and compete to innovate and evolve together, to adapt to their environment.

Since its inception, the ecosystem concept, in the field of strategy has started underlining the idea of interdependence between each single species within the ecosystem. The future of each player is indeed related to that of the others. Ecosystems are characterized by both symbiotic and antagonistic relationships, without which each single player would lose its own individual meaning. While the boundaries of an ecosystem may be blurred, companies should try to identify the players upon which their success depend [Adner and Kapoor (2009), Gawer (2009, 2021), Gawer and Cusumano (2014)]. The ecosystem also focuses on questions of access and openness, highlighting measures such as the number of partners, network density, and actors' centrality in larger networks.

In the second stage, the banking industry is going to evolve towards platform-based ecosystems, through organizing the contributions of multiple companies that collaborate to create a unique value proposition within a thematic customer journey.

Ecosystem members must coordinate to create a unique value proposition for the consumer, which would not exist without an underlying ecosystem. The unique value proposition will offer customers new experiences, so that every participant in the experience network will be under the same umbrella name platform-based ecosystem and works towards creating value as well as competing in value extraction. This results in constant tension in the strategy development process. At this point, the balance between collaborating and competing is delicate and crucial, and requires a high degree of transparency for effective collaboration and value co-creation in order to achieve a win-win strategy in co-extracting economic value.

In comparison to platforms-based ecosystems, platforms are simple business models. Both within the B2B and B2C sectors, success will be dependent on the ability to sustain large scale investment, often over a period [(Shipilov and Burelli (2020)].

The emergence of the platform-based ecosystem can be attributed to the fact that it ultimately serves the purpose of facilitating innovation and enhancing value proposition to end customers by making innovation co-evolving [Adner and Kapoor (2010)]. It also provides impetus [Brass et al. (2004)] for interorganizational ecosystem collaboration to reduce costs and increase economies of scale and scope. These findings are also in line with what was postulated by Chesbrough (2011) on the open innovation, and the achievement of common goals.

At this stage, what matters is the openness of this new paradigm, where every player may interact within one or more surrounding ecosystems [Omarini (2018)]. All this demands a new vision that is both focused and broad, highly dynamic, and interconnectable to new value propositions, based on relationships, platforms, and the sharing of information. Under these circumstances, there is a strong need for balancing the opportunities for openness with the need for consumers' protection, which is fundamental to maintaining trust and security in the financial services market.

## 4. WHAT SHOULD BE KEPT AND WHAT NEEDS TO CHANGE?

The question regarding what needs to change and what should be kept is derived from the changes that the financial services industry is undergoing.

While the core objectives of financial intermediation may remain the same, the methods and functions relating to those objectives are changing with digital technology and market developments (namely platforms and platform-business ecosystems). Within the new environment, which is affected by so many unknown variables, it becomes important to recognize the need to change patterns of analysis.

It is time to recognize that it is difficult to adopt deterministic models of input-output. This is because organizations are cognitive systems; hence, it is important to recognize their dynamics linked to learning processes and logic transformation, especially when banking and financial services are becoming increasingly customer knowledge driven.

It is time for every organization to counteract the excesses of macroeconomic theory, which has long considered banking as a "black box", designed to mediate cash flows and income-oriented balance conditions at the global level, paying less attention to its counterparties. It is also the time to be less influenced by models based on discounted cash flows.

The new frameworks provide us with two points of analyses worth outlining. The first is that similar to their traditional counterparts, new financial services providers aspire to develop the core purposes of financial intermediation, albeit with new methods and functionaries. The second point is that in many cases there is still a banking organization or a consolidated infrastructure somewhere in the fintech and bigtech stack; similar to third-party app developers who rely on smartphone sensors, processors, and interfaces. For instance, fintech developers need banks somewhere in the stack for such things as access to consumer deposits or related account data, payment infrastructures, credit origination, and compliance management.

Although there is a new generation of banking strategies entering the market, we believe that the fundamental principles of managing each of the vertical businesses, where the new financial services providers have started entering the market (payments, lending, financial advice, etc.), keep their relevance. They are still relevant for both maintaining old equilibriums and developing new ones by improving resiliency, as well as keeping the entire industry safe and stable, albeit under different emerging frameworks.

It should be noted that the new-bank-like organizations, which are tech-driven firms, are ultimately offering financial services, and, in doing so, are all working in related businesses; hence, some of the critical industry specific issues will remain in the market.

Given all of these facts, one must bear in mind that despite the role that fintechs, or any other new financial services providers, play, we must recognize that the business of banking is still complex. This complexity has been exacerbated by digital technologies and new frameworks [Omarini (2019)].

Every third-party financial services provider must be aware of the business they are in and recognize that being part of the financial services industry is only the first layer of complexity that they need to manage. Add to that the issue of deciding whether or not they want to be part of one or more platforms or ecosystems.

Finally, it is worth remembering that banking is a people business [Omarini (2015, 2019)], which means that factors such as trust, distinct professional knowledge, soundness, and a strong culture of fact-based decision making, will remain relevant.

At this stage, the main challenge for financial services providers will be to move away from being a provider in the service of customers to becoming the customers' provider. To achieve this goal, each organization must recognize that the blurring of the lines in the industry is causing a rethinking of the definition of what banking, banks, and bank-like companies are going to be for individuals in the near future.

## 5. CONCLUSIONS

The discussions above demonstrate that banking, as a business, is not in search of relevance, but has instead started renewing itself and becoming reactive to customers' behaviors and changing habits.

Embedded finance, which seems to be one of the future trends in banking, has opened the way for an ever-infinite intermediation, because both banks and non-banks are becoming increasingly crucial to everyday life. Customers are taking a more active interest in saving and investing, as well as lending and borrowing.

As we move further into the realm of digital banking and finance, there will be a greater need for the industry to rethink a number of its old concepts, including asking what is banking and what is the role of banks. It also drives the industry to accept that money, which is the "good" exchanged in the industry, is becoming increasingly digital. Its virtuality will call on regulators and the industry to give digital money more attention, and help bridge any trust gaps that could emerge in the changing market landscape. Digital money will be increasingly demanded as attitudes and behaviors change, and will influence how value is exchanged in the future.

On one hand, there are banks that have been leading the industry for a long time and need to decide whether they adapt themselves to the many changes the industry is undergoing or being the changer. On the other hand, there are other financial services providers (namely fintechs, bigtechs, etc.), which are looking to build trust among their counterparties, be they individuals or organizations.

We want to underline that today's markets are driven by choice, and customers have an abundance of options to choose from. Hence, each business must adopt a holistic mindset and bear in mind that in the digital age every business is in a permanent state of being in the business of being chosen.

Choosing from multiple options is always based on differences, be they implicit or explicit, so that differentiation is needed to give the customer a reason to choose a particular service and related experience. Hence, differentiation is becoming one of the most important and challenging drivers for competing in the market; and at present, it is not discretionary.

This is because, in the future, a single bank or financial services provider will not necessarily be called upon to provide many more services by itself, but is expected to help customers make better use of their services and to cross-buy services from a platform-business ecosystem.

## REFERENCES


Accenture, 2018a, "The brave new world of open banking," https://tinyurl.com/bzfaf3c3

Accenture, 2018b, "Competing in the new era," report

Accenture, 2019, "Competing with banking ecosystems," https://tinyurl.com/mpmjj49v

Accenture, 2020, "Banking in 2020. New trends to watch," report

Accenture, 2021, "The future of banking," report

Adner, R., and R. Kapoor, 2010, "Value creation in innovation ecosystems: how the structure of technological interdependence affects firm performance in new technology generations," Strategic Management Journal 31:3, 306-333

Alijani, S., and R. Wintjes, 2017, "Interplay between technological and social innovation," SIMPACT working paper series 2017:3, https://tinyurl.com/2p9etpsm

Anshari M., Nabil Almunawar M, Ariff Lim S., Al-Mudimigh A., 2019, Customer relationship management and big data enabled: personalization & customization of services, in Applied Computing and Informatics, Volume 15, Issue 2, July, 94-101.

AT Kearney, 2021, "Retail banking radar: change looms in Europe," report

Bareisis, Z., 2013, "The rise of the new bank account? The quest for transactional account primacy," Celent report

Bask, A., M. Lipponen, M. Rajahonka, and M. Tinnilä, 2011, "Framework for modularity and customization: service perspective," Journal of Business & Industrial Marketing 26:5, 306-319

BIS, 2019, "Report on open banking and application programming interfaces," Bank for International Settlements, https://tinyurl.com/37cx4ana

Bleier A., Keyser A., Verleye K., 2018, "Customer Engagement Through Personalization and Customization" in Customer Engagement Marketing, 2018

Brass, D. J., J. Galaskiewicz, H. R. Greve, and W. Tsai, 2004, "Taking stock of networks and organizations: a multilevel perspective," Academy of Management Journal 47:6, 795-817

Breidbach, C. F., R. Brodie, and L. Hollebeek, 2014, "Beyond virtuality: from engagement platforms to engagement ecosystems," Managing Service Quality 24:6, 592-611

Capgemini, 2019, "World retail banking report 2019," https://tinyurl.com/2pvmscyf

Capgemini, 2020, "World fintech report," https://tinyurl.com/ysdahc2f

Capgemini, 2021, "World retail banking report 2021," https://tinyurl.com/4vcayht9

Chesbrough, H., 2011, Open services innovation: rethinking your business to grow and compete in a new era, John Wiley & Sons

Cortet, M., T. Rijks, and S. Nijland, 2016, "PSD2: The digital transformation accelerator for banks," Journal of Payments Strategy & Systems 10:1, 13-27

De Blok, C., B. Meijboom, K. Luijkx, J. Schols, and R. Schroeder, 2014, "Interfaces in service modularity: a typology developed in modular health care provision," Journal of Operations Management 32:4, 175-189

Deloitte, 2017, "Open banking. How to flourish in an uncertain future," report

Deloitte, 2020, "Banking and capital markets outlook: fortifying the core for the next wave of disruption," report

Deloitte, 2020a, "Banking on the future: vision 2020," report

Deloitte, 2020b, "Open banking around the world: towards a cross-industry data sharing ecosystem," report

Dratva, R., 2020, "Is open banking driving the financial industry towards a true electronic market?" Electronic Markets 30:1, 65-67

ECB, 1999, "The effects of technology on the EU banking systems," European Central Bank, https://tinyurl.com/4vn4mf5j

ECB, 2023, "Take-aways from the horizontal assessment of the survey on digital transformation and the use of fintech," European Central Bank, February

EY, 2017, "Fintech adoption index 2017," https://tinyurl.com/3dsw5djh

Financial Stability Board, 2017, "Financial stability implications from fintech," https://tinyurl.com/224anzud

Finastra, 2022, "Banking as a Service: Outlook 2022. Paving the way for Embedded Finance. Market assessment and imperatives for success", report

Gawer, A., 2009, "Platform dynamics and strategies: from products to services," Platforms, Markets and Innovation 45, 57

Gawer, A., 2021, "Digital platforms and ecosystems: remarks on the dominant organizational forms of the digital age," Innovation: Organization & Management 24:1, 110-124

Gawer, A., and M. A. Cusumano, 2014, "Industry platforms and ecosystem innovation," Journal of Product Innovation Management 31:3, 417-433

Iansiti, M., and R. Levien, 2004, "Strategy as ecology," Harvard Business Review 82:3, 68-78

IMF and World Bank, 2019, "Fintech: The experience so far," IMF policy paper no.19/024, https://tinyurl.com/yvhcd644

KPMG, 2016, "The profitability of EU banks: hard work or a lost cause?" https://tinyurl.com/2p82nzmt

KPMG, 2020, "Can fintech lead innovation post COVID-19?" https://tinyurl.com/4ckydb8z

KPMG, 2023, "The pulse of fintech", report, https://assets.kpmg.com/content/dam/kpmg/au/pdf/2023/the-pulse-of-fintech-h1-2023.pdf

Llewellyn, D. T., 1999, The new economics of banking (no. 5), SUERF studies

McKinsey, 2017, "Data sharing and open banking," https://tinyurl.com/4cfhbjz2

McKinsey, 2022, "Banking is radically transforming. Many banks can thrive by fundamentally changing the way that financial services are embedded into daily life", https://tinyurl.com/38jydz3p

Microsoft, Avanade, Accenture, 2017, "PSD2 and open banking: using regulation to kick-start the transformation of banking," report

Moon, S. K., J. Shu, T. W. Simpson, and S. R. Kumara, 2010, "A module-based service model for mass customization: service family design," IIE Transactions 43:3, 153-163

Moore, J. F., 1993, "Predators and prey: a new ecology of competition," Harvard Business Review 71:3, 75-86

Omarini A., 2015, Retail banking. Business transformation and competitive strategies for the future, Palgrave Macmillan

Omarini A., 2019, Banks and banking: digital transformation and the hype of fintech. Business impacts, new frameworks and managerial implications, McGrawHill

Omarini A., 2022, "Unbundling and re-bundling the open industry of banking," in K.T. Liaw (ed.), The Routledge handbook of fintech, Routledge

Prahalad, C. K., and V. Ramaswamy, 2004, The future of competition: co-creating unique value with customers, Harvard Business Press

PwC, 2018a, "Open banking and so what?" https://tinyurl.com/4ej34wr4

PwC, 2018b, "Five ingredients to catch up open banking," https://tinyurl.com/42n5jsbd

Shipilov A., and F. Burelli, 2020, "Don't confuse platforms with ecosystems," INSEAD Knowledge, https://tinyurl.com/37ruzyju

Silvestro, R., and P. Lustrato, 2015, "Exploring the "mid office" concept as an enabler of mass customization in services," International Journal of Operations & Production Management 35:6, 866–894

Tuunanen, T., A. Bask, and H. Merisalo-Rantanen, 2012, "Typology for modular service design: review of literature," International Journal of Service Science, Management, Engineering, and Technology (IJSSMET) 3:3, 99-112

Zachariadis, M., and P. Ozcan, 2016, "The API economy and digital transformation in financial services: the case of open banking," SWIFT Institute Working Paper

# ORGANIZATIONAL

# THE TRUTH BEHIND ARTIFICIAL INTELLIGENCE: ILLUSTRATED BY DESIGNING AN INVESTMENT ADVICE SOLUTION

**CLAUDE DIDERICH** | Managing Director, innovate.d[1]

## ABSTRACT

Artificial intelligence can be considered one of those technologies, like 5G, 3D printing, and virtual reality, that can disrupt the business world. While AI has the potential to solve meaningful business problems, implementing it in a way that creates value is challenging. Unfortunately, many AI proponents lack the necessary computer science and mathematics machine learning skills required for developing AI systems that pass the Turing test. This paper presents an assessment of the characteristics of AI, allowing the reader to understand what specific business problems it can solve, and describes how an AI-supported investment advice solution for wealthy private clients can successfully deliver value. By reviewing the lessons learned, I conclude that the future of AI is bright if the focus is put on applying it to those challenges that it is best suited to solve.

## 1. INTRODUCTION

Artificial intelligence (AI) can be considered as the most disruptive technology dominating the 21st century [Girasa (2020), Roubini (2022)]. While the concept of AI may sound frightening, it is impossible to ignore the value that it offers society. Semi-autonomous cars (e.g., Tesla), spoken language recognition (e.g., Siri, Alexa), purchase recommendations (e.g., Amazon, Netflix), subject tracking in photo and video cameras (e.g., in the latest products by Canon, Nikon, and Sony), as well as chatbots (e.g., Bard, BioGPT, and ChatGPT), are just some examples of solutions that rely heavily on AI. In addition, though not widely recognized, AI supports numerous non-end-user facing activities, like detecting possible credit card frauds, pricing insurance risks, or constructing investment portfolios.

Undoubtedly, AI has already destroyed certain jobs and will continue to do so in the future. However, it will also lead to the creation of new jobs, help humans focus on more creative activities and improve their lives. It will allow firms to increase efficiency, resulting in lower costs and prices, extend product customization, allow customers to solve their problems in better ways, and extract meaningful data patterns from big, complex datasets, supporting superior decision making. AI offers customers a vast treasure of value-creation capabilities and value-appropriation opportunities for organizations [Kaartemo and Helkkula (2018), Wodecki (2018)].

Despite the aforementioned potential benefits, human judgment will continue to be needed to identify those challenges that can best be solved by using AI and to design the respective solutions. Human judgment is required to determine which problems AI can and cannot solve and what data AI needs to learn and create new insights from, how AI can supplement human intelligence and where it can replace it, as well as addressing the ethical challenges associated with relegating recommendations and decision making to “digital humans” [(Diderich (1993), Chancellor (2023)].

[1] The author acknowledges the valuable feedback from Esther Gelle on an earlier version of this paper.

## 1.1 Key moments in the history of AI

Taking a short excursion into the history of AI helps us better understand AI and its value to society. While AI may be seen as a recent phenomenon, it finds its roots in the 1950s. British polymath Alan Turing first suggested that if humans can use information and reasoning to solve problems and make decisions, computers should be also able to do so. This led him to formulate the famous Turing test. Originally called the imitation game [Turing (1950)], it assesses a machine's ability to exhibit intelligent behavior equivalent to, and indistinguishable from, that of a human. As of today, no general-purpose AI system has passed the Turing test.

AI started flourishing in the late 1960s as computers became more accessible. Early work by Feldman, Feigenbaum, Minsky, Newell, Simon, Weizenbaum, and Winograd [Barr and Feigenbaum (1981)] showed promising results in applying goal-based problem solving using expert systems. Expert systems [Puppe (1993)] focus on encoding rules of human thinking into computer programs. They look literally like sophisticated "if-then-else" programs. Specialized computer languages, like Prolog and Lisp, were developed to support encoding human decision rules efficiently and effectively. Designed to focus on specific problems, expert systems were high-performing, transparent, very reliable, and offered easy-to-understand results. They were well-suited for targeted problem solving. Some of the most prominent expert systems in medicine were MYCIN (diagnosing and treating infectious diseases), DENDRAL (molecular structure prediction in chemical analysis), and CaDet (detecting cancer in early stages). All these systems are based on modeling how humans understand their decision-making process rather than how human brains work. Their main drawbacks are that they lack generality and require extensive maintenance work to update the rules.

In the 1980s, AI got a second lease of life when computing power allowed for general-purpose "artificial neural networks" (ANNs) to be trained and deployed at scale. ANNs are based on the structural understanding of the human brain rather than on encoding human decision making. Haykin (1999) describes a neural network as a "massively parallel distributed processor made up of simple processing units – the neurons, which has a natural propensity for storing experiential knowledge and making it available for use. Knowledge is acquired by the network from its environment through a learning process. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge." Its simplest form is the perceptron, a linear classifier [Rosenblatt (1958)]. ANNs are generic, problem-independent functions that use previously learned insights to calculate solutions based on input data. The insights ANNs use are derived using deep learning algorithms on typically large sets of historical problem-solution data rather than hard-coded algorithms. Work by, among others, Fukushima, Grossberg, Hopfield, Kohonen, Linkster, Peral, Schmidhuber, Rosenblatt, Vapnik, Waibel, and Werbos, substantially advanced the field of knowledge in machine learning, which stands at the center of today's AI systems [Haykin (1999)].

An important, but often forgotten, result by Amaldi and Kahn (1995; 1998) and Engel (2001) finds that optimally training a single perceptron, the simplest possible ANN, is an NP-hard problem.[2] They show that it is computationally intractable (unless using quantum computers) to train an ANN to always produce the correct solution. Moreover, it is even impossible for an ANN to be trained in a way that it finds the best possible solution within a given degree of certainty; that is, in a probabilistic sense.

Few advancements were made in AI during the 1990s. However, in the new millennium, research in machine learning got a new boost with the advent of high-speed computers, low-latency networks, and massive storage capacities, making collecting and processing large problem-solution datasets more tractable [Hwang (2018)]. Furthermore, the non-scientific community also became interested in AI, notably due to landmark achievements, such as IBM's DeepBlue beating Kasparov in chess in 1997 [Campbell et al. (2002)], Google's self-driving car experiment since 2009 [Teoh and Kidd (2017)], the Watson computer system winning the first prize on the "Jeopardy" quiz show in 2011 [Baker (2011)], Apple's voice recognition Siri (2011) and Amazon's Alexa (2014), AlphaGo winning multiple games against Go champions since 2015 [Pumperla and Ferguson (2019)], and more recently DeepL (2017) and ChatGPT (2023) relying on generative AI methods and making large language model technology [Zhao et al. (2023)] available on mobile devices.

[2] A problem is called "NP-hard" if it is suspected that no algorithm using polynomial time (versus exponential time) exists that can solve it.

## 2. UNDERSTANDING THE "INTELLIGENCE" IN ARTIFICIAL INTELLIGENCE

Consider the following business idea: an entrepreneur wants to offer customized pizzas in any shape or form, like, for example, a heart shape for Valentine's Day, a steamboat pizza for celebrating a child's birthday, or a tennis racket in honor of the 100th anniversary of the Italian Open. One challenge the entrepreneur faces is estimating the amount of tomato purée needed for each pizza. I call this problem the tomato purée pizza challenge or TPPC.

The traditional approach for solving the TPPC would be calculating the amount of tomato purée using a distinct formula for each shape. For example, $f_{\text{square}}(d) = 0.2d^2$ for a square pizza and $f_{\text{heart}}(d) = 0.2(1 + \pi/4)\ d^2/2$ as an approximation for a heart-shaped pizza. While the result of this approach is exact, transparent, and quick to calculate, it lacks flexibility. A new formula has to be developed and encoded for each new pizza form. Intelligence is associated with the different formulas $f_x()$ developed.

Addressing the TPPC using AI takes a different route. First, a large number of different pizza shapes are designed (and recorded as images). The amount of tomato purée required for each is measured empirically. This leads to an extensive problem-solution dataset *D* of pizza image-tomato purée pairs. Next, a generic ANN is trained using a supervised learning algorithm, as found in standard AI algorithm libraries, on the dataset *D*. Insights *I* are derived through learning from the dataset $D : I = \text{ANN}_{\text{learning}}(D)$. Finally, for a given image *p* of a pizza shape, the trained ANN calculates the amount of tomato purée needed using the generic function ANN(*I, p)*. Intelligence is associated with the insights *I* dynamically learned from the data rather than a hard-coded formula.

In contrast with analytical approaches, the AI solution works for any pizza shape rather than only a pre-coded subset. However, it will only deliver reasonably correct results if the ANN has been trained using a representative and large dataset of pizza image-tomato purée pairs. Furthermore, the calculated amount of tomato purée required may be way off for some pizza shapes.

This paper aims to understand when AI is an appropriate tool for solving a problem and when other methods are more suited. Different problems require different solution approaches.[3] For some problems, AI will be the most appropriate approach; for others, different solutions will prevail. Consequently, when deciding whether to rely on AI to compute a solution for a given problem, it is important to:

1) Understand what the "exact problem" that needs to be solved is.
2) Know what "historical data" is available and can be legally used for learning and insights generation.
3) Know the "value and limitations" of using possible AI solutions.

## 3. DETERMINING THE SUITABILITY OF AI FOR SOLVING A SPECIFIC PROBLEM

While AI can be used to approach many wicked problems,[4] as solving the TPPC has shown, it is by no means applicable to solving every problem. The universe of problems most suitable for AI can be classified into two categories:

1) **Pattern-matching problems:** problems in this category are solved by identifying complex structures or patterns in datasets and associating them with specific solutions. An example of a typical pattern-matching problem is image recognition, e.g., identifying a cat or a human crossing a street in a picture. Speech recognition, matching spoken waves to words, is another such problem. Playing games like Go or chess can also be handled using AI algorithms designed for pattern matching. Recently, chatbots like Bert, ChatGPT, or Galactica have used large language model algorithms to solve generic pattern-matching problems, matching chat questions to learned text. Problems in the pattern matching category are best addressed using "supervised" learning algorithms [Jo (2022a)] applied to labeled datasets. The term supervised relates to the requirement that the training dataset includes labels representing known solutions to specified problems.
2) **Classification problems:** the second category of problems well suited for AI algorithms are solved by classifying data based on unknown attributes. AI can address typical classification problems: customer segmentation, anomaly detection, or product recommendations. Unsupervised learning algorithms are typically used to address classification problems [Jo (2022b)]. In contrast with analytical approaches to solving

[3] Note for the sake of completeness that in the case of the TPPC, analytical algorithms exist, for example, using triangulation, which can approximate the surface of a generic shape without relying on AI.
[4] A wicked problem is a problem that is difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize.

classification problems, these algorithms do not need to know a priori what attributes are relevant for classifying the data. During unsupervised learning, the relevant attributes are determined implicitly and often remain hidden from the outside world. Unlike supervised learning algorithms, unsupervised learning does not require solution data; that is, training data can remain unlabeled.

Identifying what value can be created by, and appropriated from, using AI as a problem-solving approach is critical. AI has the potential to deliver significant value in two business areas:

1) Identifying patterns or attributes that are "too complex" or "take too much time for humans to identify", especially because of their multi-dimensional nature or the size of the dataset. Typical problems in this category are constructing investment portfolios, detecting credit card fraud, or tracking image data.

2) "Performing repetitive tasks", where AI is significantly faster and/or cheaper than human resources. Typical problems in this category are voice recognition, text translation, writing draft documents, or searching for specific data items.

### 3.1 Five premises for using AI-based problem solving

Not all problems are sound for AI solving. Five premises must be satisfied to solve a wicked problem using AI successfully. These are:

(1) The problem at hand "cannot be solved using analytical algorithms", or using analytical algorithms is computationally infeasible, although theoretically possible.

(2) The problem "can be solved by relying on available historical data". The solution is not entirely novel. This does not mean that existing data must include the solution but that it can be reasonably inferred from it.

(3) There exists appropriate "labeled" or "unlabeled datasets" (depending on the type of problem) that can be legally used for AI learning purposes.

(4) Relying on a suboptimal or incorrect solution "is a viable option".

(5) The problem solver is "not faced with any moral hazard due to an incorrect solution" computed by AI. No human lives are at risk if AI fails to find the right solution.

Premise (1) states that a problem-specific algorithm exhibiting validated properties is preferred to a problem-agnostic machine learning approach. Although this may seem obvious, it means that AI should not be used as a replacement for human domain-specific knowledge. While AI allows for combining existing knowledge in a way that humans might not have thought of, premise (2) states that AI cannot invent new knowledge. With efficient data-collecting resources available, satisfying premise (3) should be straightforward. However, it is not. Legally collecting high-quality data often poses an insurmountable challenge. Finally, premises (4) and (5) address the challenge that AI cannot guarantee the correctness of its results. In many situations, AI cannot even offer the reasoning that has led to the solution, thus making the work of human result validation tedious, if not impossible. Recent research in explainable AI [Holzinger et al. (2022)] focuses on addressing that challenge. Premise (5) stipulates that if the use of AI could lead to moral hazard, it must be used primarily as a decision-support tool complemented by human expertise and/or analytical algorithms.

### 3.2 Challenges faced by AI

Applied to the right problems, AI can offer solutions humans could not think of. However, these solutions have some caveats that must be understood before relying on them.

Whether relying on labeled or unlabeled data for learning, generic AI algorithms make it possible to find correlations in the training data, but not causalities. Pearl (2000) and Pearl and Mackenzie (2018) have shown that "data alone can never answer causal questions. They [AI algorithm developers] require to formulate a model of the process that generates the data or at least some aspects of that process." Incorrectly assuming causality when only a correlation exists is one of the biggest mistakes one can make when relying exclusively on data to solve problems. This is no different for AI. Many, if not all, sophisticated AI algorithms include some sort of domain-specific model to support deriving causalities. For example, ChatGPT has learned that most famous sports journalists have covered the Olympic games. However, when ChatGPT is asked what events a known sports journalist has covered, it incorrectly infers the causality that such a journalist must have covered, with a high probability, the Olympic games, although only a correlation exists.

Training a generic ANN in such a way that it correctly classifies the largest possible number of data elements is a computationally intractable problem. This means that unless using domain-specific modeling when designing and training an ANN, it is impossible to ensure, even in-sample, the quality of any result. It is computationally infeasible to train a generic ANN in such a way that it correctly solves a given percentage of problem instances; that is, offers a probabilistically correct answer.

A third, and even more vital, challenge that many AI systems face is that they are black boxes. AI typically provides a possible solution but cannot explain how that solution was derived. For example, an ANN used for recognizing animals in images was trained using, among others, horse images that included a copyright notice (which non-horse images did not have). When using the trained ANN on new images, it incorrectly identified any image containing a copyright notice as an image of a horse [Lapuschkin et al. (2019)]. Although research in designing explainable AI (XAI) algorithms has made progress in recent years [Samek et al. (2017)], notably by attributing the statistical probabilities of each input to the result component, there is still a long way to go to come up with domain-independent AI algorithms that offer explainable solutions. Most promising research in XAI focuses on designing interpretable models using decision trees, Bayesian networks, and sparse linear models [Rudin (2019)].

## 4. TOWARDS MORE SUCCESSFUL INVESTMENT ADVICE: AN AI CASE STUDY[5]

Offering "customized investment advice" (CIA) as a paid service has become one of the most prominent offerings in private banking. One of the reasons for this is tighter regulations imposed on investment advisors to avoid conflicts of interest. Another is that customers seek help navigating the ever more complex investment universe without delegating the final investment decision. Private banks like CIA because it can be sold in a way to generate recurring revenues.

### 4.1 Understanding CIA

A naïve manager would consider CIA a product recommendation problem, similar to Amazon suggesting to its customers which books to buy based on past purchases or Netflix proposing what movie to watch next based on learned user preferences. Unfortunately, advising CIA customers is more complex, as it involves multiple stakeholders with different goals and preferences: the investor as the customer, the investment advisor and their employer as service providers, and the investment product providers as the manufacturers.

Investors look for investment recommendations that meet their risk profiles, reflect their market expectations, and bring them closer to their financial goals. Investment advisors aspire to advise clients effectively, maximizing the probability that the investor will act upon their advice and be happy. They also look for help navigating the ever-growing universe of investment products, each with its features and caveats. Their employer, on the other hand, wants to maximize value capturing. Finally, investment product providers look for their offerings to be recommended by the investment advisor. Based on these observations, the CIA service can be reformulated as a decision problem suitable for solving using AI.

### 4.2 Formulating the CIA as an AI problem

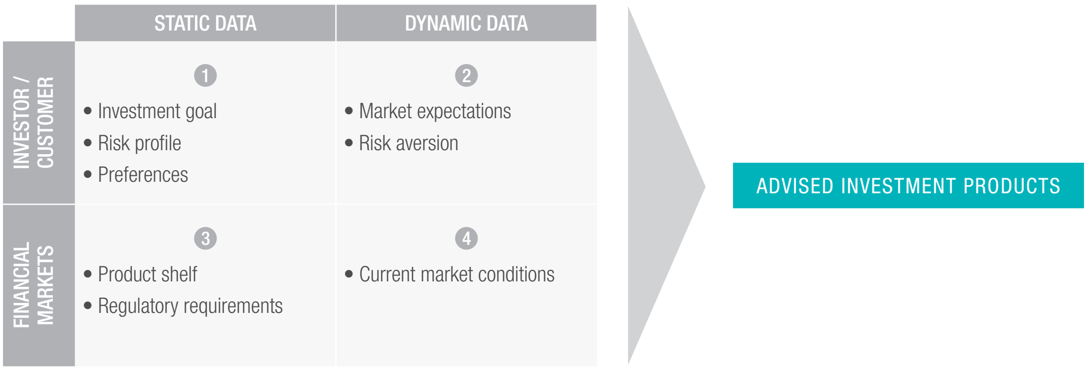
To determine which products to recommend to its client, the investment advisor must evaluate the function shown in equation (1), where the parameters ❶, ❷, ❸, and ❹ are defined in Figure 1 as the data universe used in offering CIA.

$$r(\text{❶}, \text{❹}, a(\text{❶}, \text{❷}, \text{❸}, \text{❹})) = \{\text{set of advised investment products}\} \qquad (1)$$

Function $r()$ is an analytical function encoding applicable regulations ensuring that investment recommendations align with them. The $a()$ function computes the set of products from the product shelf that the investment manager recommends to the investor based on their investment goal, risk profile, subjective market views, preferences, and given objective market conditions. While function $r()$ fails the AI premises (1), (4), and (5), defined in Section 3.1, function $a()$ meets all five AI premises. As such, function $a()$ is well suited to be implemented using a multi-layer feed-forward ANN.

[5] Note that the case study presents a high-level description of how AI can be used to improve the customized investment advice service. For the sake of readability, the description has been simplified. Details, including some modeling and intermediary data processing steps, have been omitted.

**Figure 1:** Data universe used in offering CIA




Notes: Static data is time-independent, whereas dynamic data changes over time.
Investor/customer data is specific for each investor, whereas financial market data is the same for all investors.

**Table 1:** AI technology stack model for implementing and operating an AI solution

| TECHNOLOGY STACK | DEFINITION | FUNCTIONALITY FOCUSING ON THE CIA PROBLEM |
|---|---|---|
| **Infrastructure platform** | Hardware underlying the AI solution | Generic AI implemented on a cloud infrastructure (i.e., AI as a service) |
| **Framework** | AI architecture | Multi-layer feed-forward ANN |
| **Learning algorithm** | Specific machine learning algorithm(s) used | Off-the-shelf, supervised learning algorithm |
| **Data pipeline** | Data source and data management platform | Proprietary client data, proprietary market and risks assessment data, public market data |
| **AI service** | Well-defined service applying the learning algorithm to the data pipeline, consistent with the framework using the infrastructure platform | General purpose API makes it possible to learn/calculate the parameters; that is, the weights associated with the nodes, of the multi-layer feed-forward ANN |
| **Scoring algorithm** | Domain-specific AI solution addressing the business problem | Custom-build capabilities resulting in investment product recommendations based on client-specific data and current financial market conditions |

Source: Based on Tsaih et al. (2023)

To develop and train an ANN that implements the function *a*(), I use a variation of the AI technology stack proposed by Tsaih et al. (2023), as shown in Table 1. A structured approach focusing on specific outcomes in a well-defined order helps avoid mixing different concepts, which could result in a sub-optimal, often even non-working, AI solution. It also helps make it easier to identify the exact problem that needs to be solved by distinguishing between technology, framework, training data, learning, and scoring. By encapsulating all historical data aspects into the data pipeline stack, the approach ensures that the appropriate data is available and can be legally used. Finally, distinguishing between learning and scoring algorithms helps identify value and limitations of the AI designed solution.

#### 4.2.1 INFRASTRUCTURE PLATFORM

As the problem to be solved is a typical, although multi-stakeholder, pattern-matching problem, there is no need for a problem-specific AI infrastructure platform. Furthermore, due to the difficulty of estimating the computing resources required for training and scoring the ANN a priori, I rely on a generic AI cloud infrastructure such as Amazon AWS AI, Microsoft Azure AI, or IBM Watson ML.

#### 4.2.2 FRAMEWORK

Next, I model the function *a*() as a multi-layer feed-forward ANN, where the input layer ingests the parameters ①, ②, ③, and ④, excluding regulatory requirements. The output layer is associated with the recommended investment products from the product shelf.

To keep the designed framework as simple as possible, I refrain from integrating back-propagation that would allow the ANN to learn by itself from the market performance of the recommended investment products while scoring and correcting faults in internal stages of the network. Instead, I regularly re-train the ANN when relevant new investor and market data becomes available.

#### 4.2.3 LEARNING ALGORITHM

While designing the third level of the AI technology stack, I use a standard supervised learning algorithm offered by the cloud infrastructure platform rather than developing a proprietary one. Such algorithms typically depend on gradient-driven optimization combined with heuristics to speed up the computations and avoid local optima.

#### 4.2.4 DATA PIPELINE

Organizing and managing the data pipeline is the most challenging part of designing, building, and implementing an ANN. Each grey vertical box in Figure 2 (training data) represents a separate dataset for training the ANN. It is specific for a given investor at a given point in time. The parameters ①, ②, ③, and ④ represent the input dataset for the point in time *t*. The "investment advice" represents the output data or label associated with the input data; that is, the portfolio holdings and investment products the investor chose at time *t* as their preferred investments.

I use raw data collected from the KYC[6] process and from risk profiling the investor, as required by regulations, as static investor data (①), describing their financial goals, risk profile, and preferences. Relying on raw data allows the ANN learning algorithm to potentially identify hidden correlations between attributes while remaining fully aware of potential noise in the collected data that could negatively impact the outcome [Kahneman et al. (2021)].

Unfortunately, investor market expectations and risk aversion data are typically unavailable at a given time *t*. Consequently, I derive the investor's expectations and risk aversion (②) from their portfolio holdings at time *t*. To do so, I associate specific market expectations and risk preferences with each investment product. For example, holding technology stocks is associated with the expectation that equity markets will grow more than GDP and have a low risk aversion, whereas holding inflation-linked bonds is associated with the investor expecting inflation to rise faster than markets expect and being risk averse by seeking protection.

The third data category represents the dynamic market data (③). It describes the observed current market conditions, like inflation rate, GDP, unemployment rate, and stock market valuations. In contrast with input ① and ②, the current market conditions data is independent of any specific investor and thus identical in all datasets for a given time *t*.

Parameter ④ represents the shelf of investment products available at time *t*. Furthermore, I assume that the regulatory requirements (part of parameter ④) are codified in an analytical function and are not derived from the dataset used for training the ANN.

To label the output or learning datasets (investment advice), I assume that the investors' portfolio holdings at any given time reflect their actual investment decision. They represent the investment products that the investment advisor should have recommended to the investor at that point in time.

#### 4.2.5 AI SERVICE

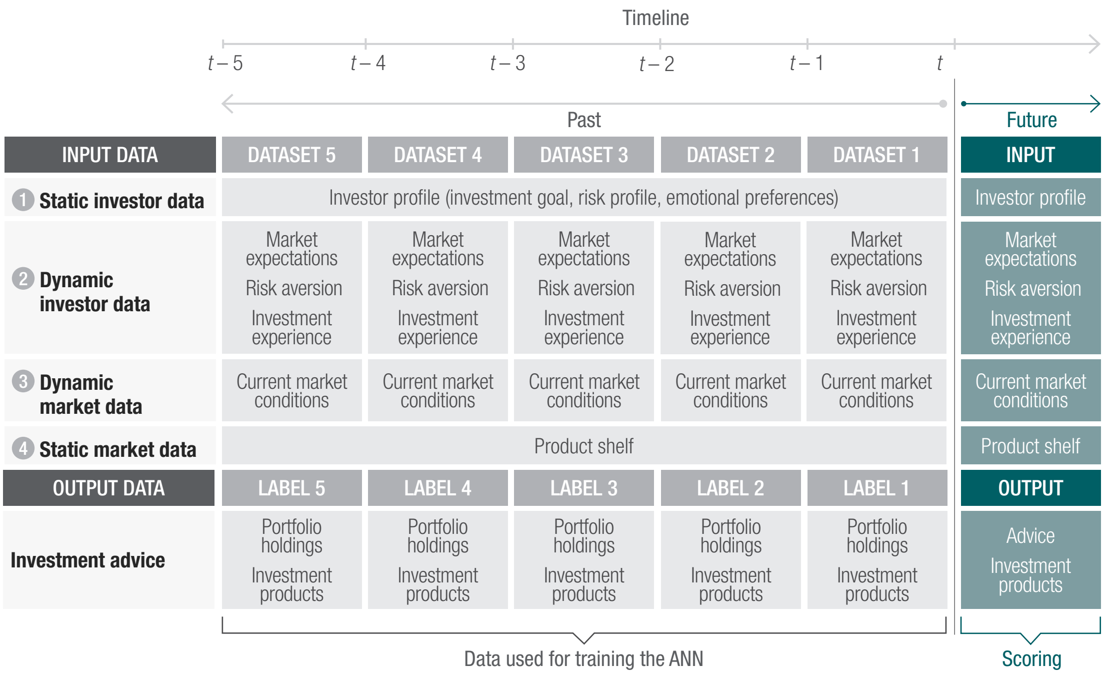
The AI service layer implements the supervised learning algorithm. It is applied to the data pipeline using the specific API the infrastructure platform provides. The outcome is a well-defined function *a*( ), which can subsequently be used to compute a set of possible investment product recommendations based on static and dynamic investor data combined with current market conditions and available product shelf.

#### 4.2.6 SCORING ALGORITHM

Finally, the scoring algorithm calculates what investment products the investment advisor should recommend to the investor based on the insights learned by the ANN. It implements evaluating equation (1) and is illustrated in Figure 2 by the green vertical box (scoring).

[6] KYC = Know your client.

**Figure 2:** Data pipeline used for learning (datasets 1 to 5) and scoring (input/output)




The static investor data (parameter 1) and the current product shelf (parameter 4) are extended by the investor formulating their expectations of the financial markets and their current risk aversion (parameter 2), i.e., their expectations of inflation, economic growth, future unemployment rates, etc. Current market conditions (parameter 3) are based on observed inflation, GDP, unemployment rate, etc. It is important to note that in this specific CIA solution only the investor makes predictions of the markets. The investment advisor's judgmental role is curtailed to describing the current market conditions.

Finally, the output from function *a*( ) is passed through the regulatory requirements filter *r*(), resulting in a set of investment products recommended to the investor that is regulatory compliant and available from the investment advisor's firm.

A point to remember is that the value of the investment advice depends on the model assumption that the investor was happy with their past investments made through their portfolio holdings and that they reflected the investor's market expectations.

### 4.3 Creating value by applying the model

The success of using AI in the designed CIA solution is based on four assumptions:

1) Similar customers (i.e., concerning investment goals, risk profile, and rationale, as well as psychological preferences) invest similarly in similar environments.

2) Investors accept that investment products advised to them may perform poorly, given their expectations.

3) Investors were happy with their past investment decisions (or the decisions that they were unhappy with were flagged as such and subsequently removed from the training dataset).

4) AI can, within reasonable boundaries, correctly identify relevant attributes in the presented datasets and classify data accordingly without requiring human modeling or manual intervention.

When these four assumptions are met, investment advice computed by AI should be expected to be superior and more consistent than guidance from human investment advisors alone. One reason is that AI has superior capabilities in identifying those attributes that matter most to investors, which may go unnoticed by human investment advisors. Another reason is that AI is able to deal better with a large universe of potential investment products than humans. Furthermore, AI does not suffer from human judgmental biases, like anchoring, availability, conjunction fallacy, optimism, loss aversion, framing, sunk costs, or overconfidence [Kahneman (2011)]. Under the assumption that the world will not be disrupted, and the future (even if non-natural and complex) will still relate to the past, using AI to support CIA should lead to a higher acceptance rate of investment advice provided and, therefore, happier customers. The potential drawback that AI cannot as yet offer explicit explanations for its advice can be mitigated by using AI as a tool to support the investment advisor's expertise rather than to replace them.

## 5. CONCLUSION

More is needed than just the artificial part of AI to successfully apply it to solving wicked problems. As with any problem-solving approach, only well-understood challenges can be successfully solved. One cannot expect AI to understand a poorly formulated problem, let alone solve it. This means that possessing big data is not enough. Implementing AI to create value for its users and allow its creators to capture part of that value requires diverse skills. Hard-core mathematical and computer science skills are too often left on the backbench or completely ignored.

### 5.1 Lessons learned from the past

Several highly relevant insights can be gained from research in AI, the presented case study, and experience implementing AI solutions to solve wicked business problems involving large datasets.

- To create and capture value in business, problem solving requires understanding the problem and identifying how a solution creates value for the stakeholders involved.
- Just because a problem involves substantial amounts of data does not make it necessarily suitable for solving using AI.
- AI is well-suited for solving problems that require identifying patterns in large datasets, which are structurally too sophisticated for the human eye to detect.
- AI best identifies correlations and correlation-like structures between data elements, especially non-linear ones, and clusters similar data elements.
- Analytical problem-solving techniques will outperform AI in most cases where computationally feasible analytical solutions to the considered problem exist.
- The most important caveat to consider when relying on AI is that it is mathematically impossible for any AI algorithm, unless combined with causality models, to guarantee the correctness of the calculated solution, even in probabilistic terms.

### 5.2 Looking into the future

While the lessons learned from past experiences with AI may sound grim, AI offers enormous opportunities when correctly applied. There are a considerable number of challenges where analytical approaches have failed or performed poorly. In situations where solving a problem requires mining large historical datasets, AI will outperform traditional algorithms in all but the most straightforward cases.

Two key challenges must be addressed to fully exploit AI and succeed at the Turing test. First, machine learning algorithms must include an explainable component, whether relying on supervised, unsupervised, or reinforcement learning. Black-box AI will not survive the scrutiny required for large-scale and/or mission-critical deployment. Second, AI must move from learning correlations to creating causal knowledge. As such, AI must allow for combining with human-designed causality models.

We are a long way from machines being genuinely creative; that is, creating knowledge that cannot be derived by combining existing knowledge. However, taking an optimistic-realistic approach to AI will make it possible to create and capture value beyond efficiency and effectiveness gains. AI is a sophisticated tool that, when used wisely, especially in combination with other tools (and humans), will allow for shaping critical aspects of our future.

## REFERENCES


Amaldi, E., and V. Kahn, 1995, "The complexity and approximability of finding maximum feasible subsystems of linear relation," Theoretical Computer Science 147:1-2, 181-210

Amaldi, E., and V. Kahn, 1998, "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems," Theoretical Computer Science 209:1-2, 237-260

Baker, S., 2011, Final jeopardy - man vs. machine and the quest to know everything, Houghton Mifflin Harcourt

Barr, A., and E. A. Feigenbaum, 1981, The handbook of artificial intelligence (Vol. I & II), Heuristech Press

Campbell, M., A. J. Hoane Jr., and F.-h. Hsu, 2002, "Deep Blue," Artificial Intelligence 134:1-2, 57-83

Chancellor, S., 2023, "Towards practices for human-centered machine learning," Communications of the ACM 66:3, 78-85

Diderich. C., 1993, "Expert systems and their risks: an ethical study," in International Conference on Computer Science, Communication, and Society: A Technical and Cultural Challenge, Neuchâtel, Switzerland

Engel, A., 2001, "Complexity of learning in artificial neural networks," Theoretical Computer Science 265:1-2, 285-306

Girasa, R., 2020, Artificial intelligence as a disruptive technology, Palgrave Macmillan

Haykin, S., 1999, Neural networks: a comprehensive foundation, Pearson Education

Holzinger, A., A. Saranti, C. Molnar, P. Piecek, and W. Samek, 2022, "Explainable AI methods – a brief overview," International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers, LNAI 13200, 13-38, Springer

Hwang, T., 2018, "Computational power and the social impact of artificial intelligence," SSRN, https://tinyurl.com/4wfytsz3

Jo, T., 2022a, "Supervised learning," in Jo, T. (ed.), Machine Learning Foundations, Springer

Jo, T., 2022b, "Unsupervised learning," in Jo, T. (ed.), Machine Learning Foundations, Springer

Kaartemo, V., and A. Helkkula, 2018, "A systematic review of artificial intelligence and robots in value co-creation: current status and future research," Journal of Creating Value 4:2, 211-228

Kahneman, D., 2011, Thinking, fast and slow, Farrar, Straus and Giroux

Kahneman, D., O. Sibony, and C. R. Sunstein, 2021, Noise. A flaw in human judgment, Little, Brown Spark

Lapuschkin, S., S. Wäldchen, A. Binder, G. Montavon, W. Samek, and K.-R. Müller, 2019, "Unmasking clever Hans predictors and assessing what machines really learn," Nature Communications 10(1096)

Pearl, J., 2000, Causality: models, reasoning and inference, Cambridge University Press

Pearl. J. and D. Mackenzie, 2018, The book of why: the new science of cause and effect, Effect. Basic Books.

Pumperla, M., and K. Ferguson, 2019, Deep learning and the game of Go, Manning

Puppe, F., 1993, Systematic introduction to expert systems, Springer Verlag

Rosenblatt, F., 1958, "The perceptron: a probabilistic model for information storage and organization in the brain," Psychological Review 65:6, 386-408

Roubini, N., 2022, Megathreats, John Murray Publishers

Rudin, C., 2019, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nature Machine Intelligence 1:5, 206-215

Samek, W., T. Wiegand, and K.-R. Müller, 2017, "Explaining artificial intelligence: understanding, visualizing and interpreting deep learning models," ITU Journal: ICT Discoveries 13 (Special issue no. 1)

Teoh, E. R., and D. G. Kidd, 2017, "Rage against the machine? Google's self-driving cars versus human drivers," Journal of Safety Research 63, 57-60

Tsaih, R.-H., H.-L. Chang, C.-C. Hsu, and D. C. Yen, 2023, "The AI Tech-Stack Model," Communications of the ACM 66:3, 69-77

Turing, A. M., 1950, "Computing machinery and intelligence," Mind LIX:236, 433-460

Wodecki, A. (2018). Value Creation and Competitive Advantage Models. In A. Wodecki, Artificial Intelligence in Value Creation (pp. 1-70). London, UK: Palgrave Macmillan

Zhao, W. X., et al. (2023). A Survey of Large Language Models, https://tinyurl.com/yjf68va5

# DUTY CALLS – BUT IS INDUSTRY PICKING UP?

**JESSICA TAYLOR** | Consultant, Capco
**IVO VLAEV** | Professor of Behavioral Science, Warwick Business School
**ANTONY ELLIOTT OBE** | Founder, The Fairbanking Foundation[1]

## ABSTRACT

The FCA's Consumer Duty regulation aims to transform financial services for customers by requiring firms to consider the needs, characteristics, and objectives of all their customers, and how they behave, at every stage of the customer journey. Its success, however, is dependent on compliance from firms and with new regulations, there often exists a policy-implementation gap whereby policies do not lead to changes in behavior. This study provides a novel approach by applying "behavioral science frameworks" to compliance with financial regulation, improving outcomes for customers under the FCA's Consumer Duty and future financial regulatory change.

## 1. INTRODUCTION

In July 2022, the U.K.'s Financial Conduct Authority (FCA) published the final version of the Consumer Duty, a standard aimed at driving a new principle requiring financial services firms in the U.K. to act to deliver good outcomes for retail customers [FCA (2022a)]. Since the financial crisis, government and regulatory authorities have sought to tackle causes of weak competition in financial services, motivated by concerns regarding poor customer outcomes, high prices, and poor value [WBF (2023)]. While competition has improved, as recognized by the FCA in a strategic review of retail banking, it is still the case that consumers, especially those in vulnerable situations, experience subpar outcomes. This issue has become more pertinent as the sector adapts to post-pandemic changes in consumer behavior and increasing advancements in technology [FCA (2022b)]. In response, the Consumer Duty requires firms to consider the needs, characteristics, and objectives of all their customers, and how they behave, at every stage of the customer journey. While acknowledging that consumers are ultimately responsible for their actions, the principle obliges firms to provide a layer of protection for customers due to the imbalance in bargaining power and expertise between them and the firms, and due to customers' susceptibility to cognitive biases, which may hinder their decision making [FCA (2022a)].

The Consumer Duty consists of three components: a new consumer principle that requires all financial services firms to "act to deliver good outcomes for retail customers," cross-cutting rules to support the new principle, and four outcomes, each with rules for firms to follow to drive these good outcomes for their customers. Besides acting to deliver good customer outcomes, firms will need to understand and demonstrate whether those outcomes are being met. The deadlines are tight, as all new and on-sale products and services must comply by July 2023, and all closed products and services by July 2024 [FCA (2022a)]. The Duty employs an innovative data-led supervisory strategy that transforms the FCA's ability to supervise. By asking firms to police themselves, the FCA is

[1] This article was written in partnership with The Fairbanking Foundation.

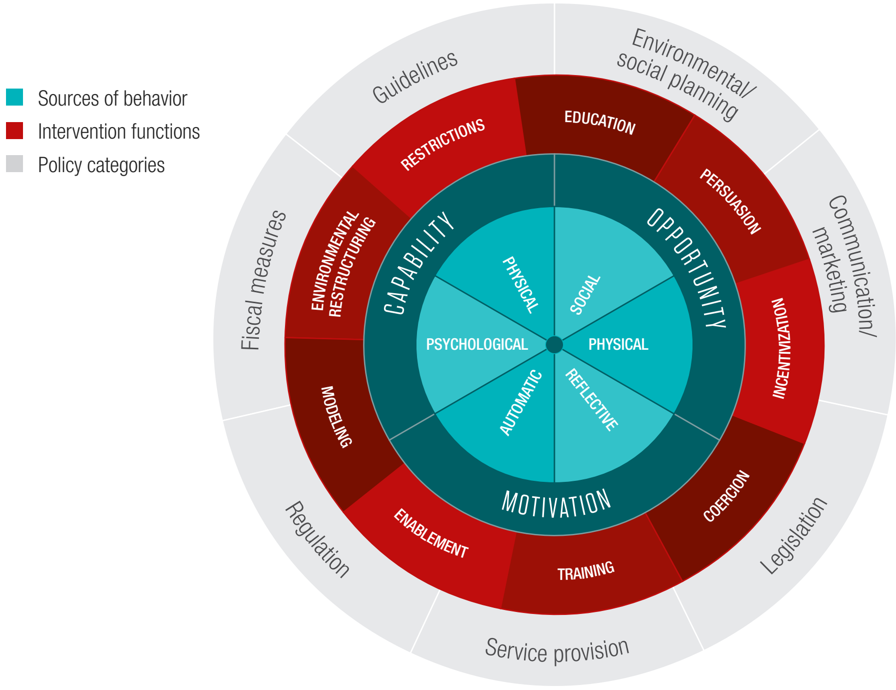
effectively putting the onus on governing bodies to determine whether a given firm is delivering good outcomes; hence enabling the FCA to apply limited supervisory resources more effectively [WBF (2023)]. This outcomes-based approach has the potential to be transformational, providing benefits for the regulator, the firms, and the consumers. Its success is, however, dependent on compliance from firms and with new research and regulations, there often exists a policy-implementation gap whereby policies do not lead to effective changes in behavior [Hudson et al. (2019)].

The literature suggests that compliance with financial regulations involves a variety of individual behaviors influenced by a combination of instrumental and normative factors, as well as cognitive and behavioral processes. Behavioral science insights have been successful in changing consumer compliance behaviors in the financial services industry. However, further research is needed to explore the application of behavioral interventions to change the behaviors of finance professionals, particularly in the context of compliance. The use of "behavior change frameworks" offers a potential avenue for designing effective interventions that address the barriers and enablers of compliance behaviors, ultimately improving customer outcomes in the context of the Consumer Duty.

## 2. BEHAVIOUR CHANGE WHEEL (BCW)

Behavior change frameworks, such as the "behavior change wheel" (BCW) [Michie et al. (2014)], are useful tools for understanding barriers and enablers of compliance behaviors and designing effective behavioral interventions. The BCW's core model, the COM-B model of behavior change, identifies capability (C), opportunity (O), and motivation (M) as the three drivers of behavior (B) (Figure 1). To influence behavior, the individual must be motivated to change, while also possessing the capability and opportunity to do so. Capability can be psychological, relating to knowledge and skills, or physical, related to physical abilities. Opportunity can be physical, such as environmental factors, or social, influenced by interactions with others. Motivation can be automatic, such as impulses and inhibitions, or reflective, related to planning and conscious decision making. The COM-B model suggests that the absence of any of these components can contribute to the policy-implementation gap and offers "behavior change techniques (BCTs) depending on which element requires modification. The efficacy of the COM-B model has been thoroughly demonstrated in the context of healthcare policies [Handley et al. (2016)], but its application to the financial services industry is relatively unexplored.

**Figure 1:** Behavior change wheel




Source: Michie et al. (2014)

## 3. PRESENT STUDY

Insights from behavioral science provide the potential to understand how and why professionals behave in response to regulation and to identify which factors contribute to variability in effectiveness and can be changed to improve outcomes [Peat (2021)]. This study investigated the policy-implementation gap in finance by using the BCW to identify barriers to behaviors that consumers and professionals identify as essential for achieving good customer outcomes, and potential interventions to target these. The findings are expected to have implications for improving the success of financial regulations in the future and ultimately lead to improved financial outcomes for customers. The research was conducted in two phases, answering the following research questions:

1. Comparing the views of consumers and professionals, what target behavior will have the largest impact on good customer outcomes?
2. Using the BCW, what are the barriers and facilitators to the target behavior, and how can these be modified through "intervention functions" and "behavior change techniques"?

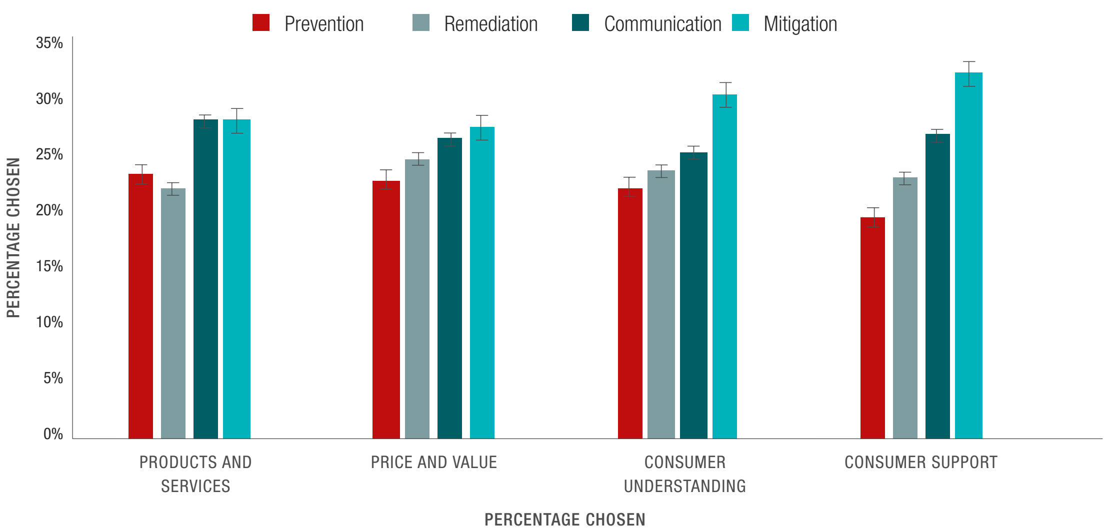
A mixed-methods study design was utilized, consisting of a quantitative survey with consumers and qualitative semi-structured interviews with practitioners. A survey was conducted with users of unsecured loan products to identify which actions they believed would lead to the most satisfactory outcomes. To supplement consumers' views, interviews were conducted with a group of financial services professionals and analyzed inductively to determine their views on the required behaviors. This was compared to the behaviors deemed important by consumers to define the target behavior that contributes to the successful implementation of the Consumer Duty standards. To understand the barriers and facilitators to this target behavior, the interviews were then analyzed deductively using the COM-B model.

## 4. RESEARCH QUESTION 1: WHAT TARGET BEHAVIOR WILL HAVE THE LARGEST IMPACT ON GOOD CUSTOMER OUTCOMES?

To gauge consumers' expectations from firms, and to define the behaviors that contribute to the successful implementation of the Consumer Duty standards, a survey was conducted with users of unsecured loan products to understand which actions, taken by the bank in response to the identification of consumer harm, they believed would lead to the most satisfactory outcomes. Interviews were then conducted with a group of financial services professionals to determine their views on required behaviors for compliance and for culture change, and to compare this to the behaviors deemed important by consumers.

To determine which actions were most favored by consumers, survey responses were ranked by the frequency by which they were chosen. To determine the drivers of this ranking, frequencies were calculated based on which outcome they

**Figure 2:** Percentage of time an action from a category was chosen, by outcome




Note: Error bars show standard errors.

**Figure 3:** Frequency of actions chosen

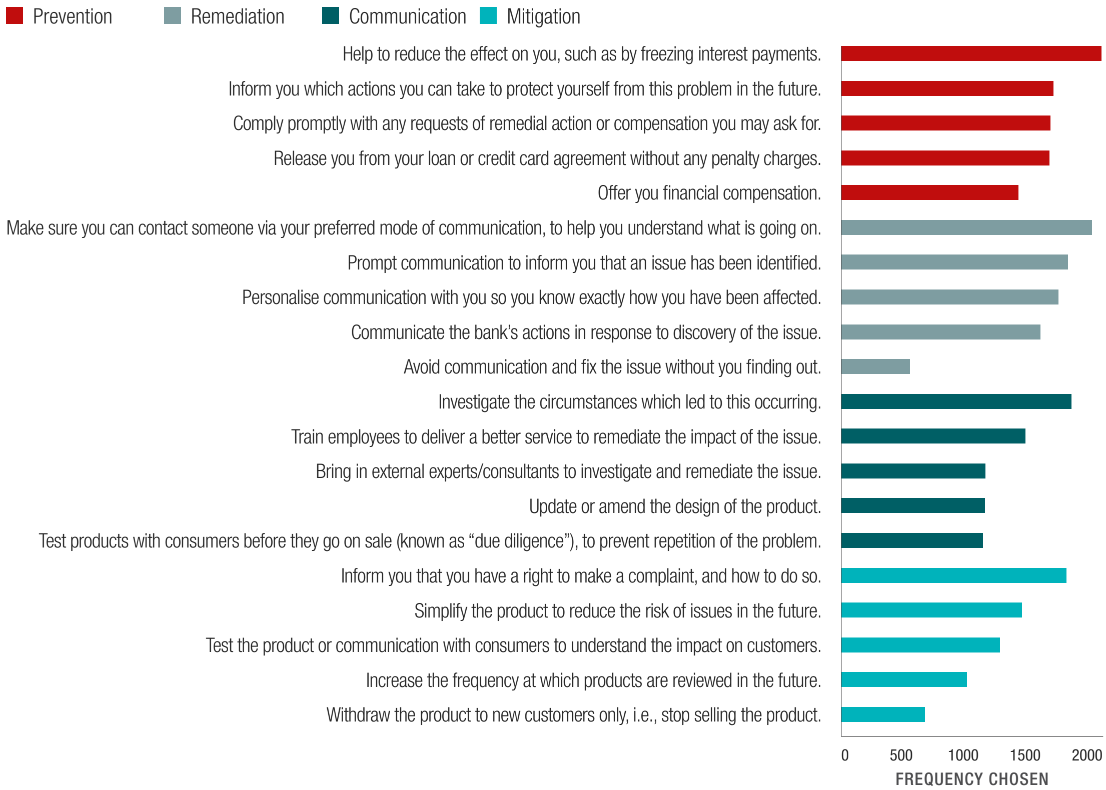
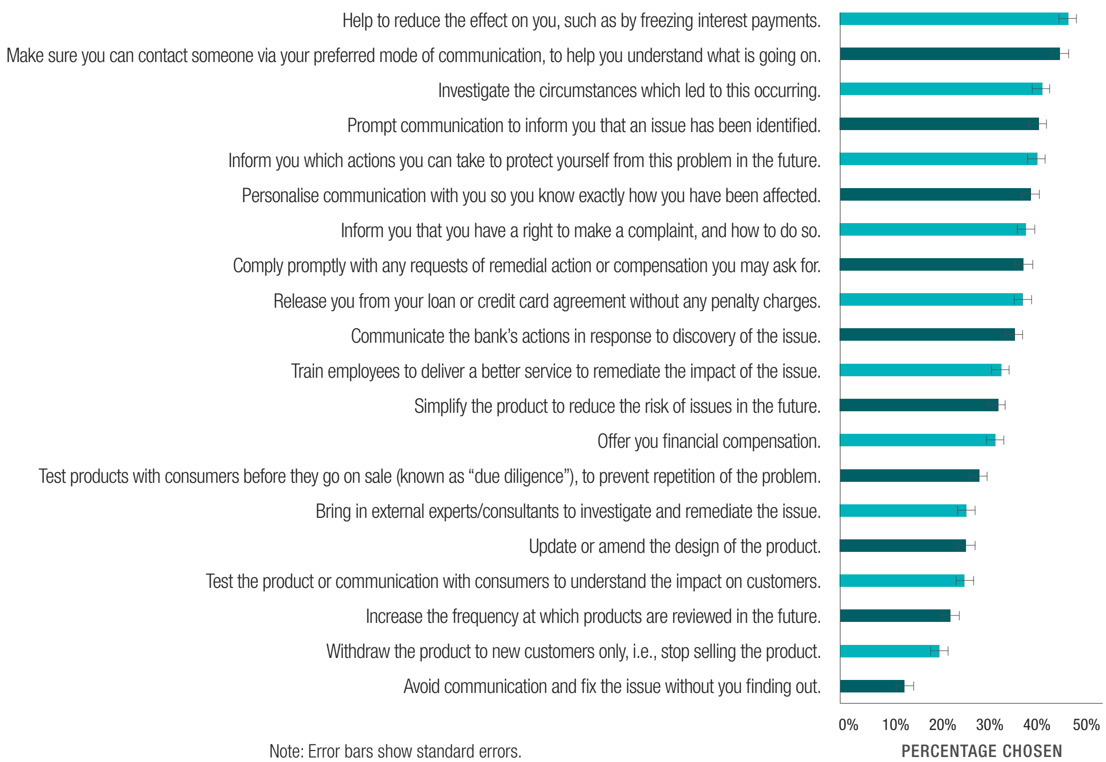



**Figure 4:** Percentage of times each action was chosen when presented as an option




Note: Error bars show standard errors.

were in response to, as well as the demographic factors of social grade, education, and household income. Chi-squared tests were performed to determine whether these had a significant effect on preferences.

To determine which behaviors practitioners partook in the most frequently, interview data were analyzed inductively using Braun and Clarke's (2006) six-phase thematic analysis process (Figure 2). Initial codes were generated by labeling each smallest possible data fragment with a descriptive unit of meaning. Codes were based on quotes that reflected a behavior performed by the practitioner, which related to either compliance or a culture change in response to the Consumer Duty. Initially, 218 codes were identified, but throughout an iterative process of condensing, synthesizing, and restructuring, codes were merged into 67 distinct behaviors [Miles et al. (2014)]. Codes were then collated into potential themes and themes were categorized deductively based on whether they contributed towards compliance – i.e., they were required in the FCA's Consumer Duty Guidance – or whether they related to a wider cultural change.

As per Braun and Clarke's (2006) process, themes were then reviewed to ensure that they fit with the coded extracts and the entire dataset. Themes were refined iteratively throughout an ongoing process to create distinctive names and definitions. To ensure reliability, an extract of the raw data, the original code(s), and the related themes, with definitions, was created for one of the interviews, which was analyzed by trustees at the Fairbanking Foundation. Any discrepancies were discussed, and a conclusion was drawn that satisfied both coders and prior coding was modified to reflect the change. To answer the research question, the behaviors identified and prioritized in consumer survey findings were compared to the behaviors identified by practitioners in interview findings.

## 4.1 Quantitative results

### 4.1.1 WITHIN EACH OUTCOME

The output of the survey is a ranking of consumers' preferences for each of the actions in comparison to each other. Mitigation was the most chosen category overall (29.1%), followed by communication (26.2%), and remediation (23.0%). Prevention was the least preferred category (21.7%). This ranking was the same across all four outcomes bar Products and Services, whereby remediation was the least chosen frequently (Figure 2). A chi-square test of independence showed a significant association between Consumer Duty outcome and preferred category of action, $X^2$ (9, N = 28,032) = 81.59, p = .000. For Consumer Support, consumers were significantly less likely to have chosen prevention ($p < .05$), and significantly more likely to have chosen mitigation ($p < .1$).

### 4.1.2 WITHIN EACH CATEGORY

The most chosen action was for banks to: "Help to reduce the effect on you, such as by freezing interest payments", which was categorized as mitigation, and the least chosen was to "avoid communication and fix the issue without you finding out", categorized as communication. See Figure 3 for the frequency of each action, by category.

### 4.1.3 ACROSS CATEGORIES

Figure 4 shows a ranking of the times each action was chosen, as a percentage of the number of times it was presented to participants.

### 4.1.4 DEMOGRAPHICS

To determine the drivers of preferences, frequencies by category were calculated based on the demographic factors of education, household income, and occupation. A chi-square test of independence showed no significant association between level of education and preferred category of action, $X^2$ (12, N = 27,640) = 13.64, p = .324; between occupation and preferred category of action, $X^2$ (9, N = 28,032) = 11.50, p = .243; or between household income and preferred category of action, $X^2$ (12, N = 28,032) = 10.31, p = .589.

### 4.1.5 SUMMARY

The most frequently chosen actions by consumers were for firms to help reduce the effect on them; make sure they can contact someone; and investigate the circumstances which lead to this occurring. The least chosen actions were for firms to avoid communication and fix the issue without you finding out; withdraw the product to new customers only; and increase the frequency at which products are reviewed in the future. In essence, consumers sought immediate harm limitation, ease of communication, and proactive investigation, with a strong preference for personally beneficial and short-term actions over those with broader implications. Prevention was the least prioritized category for consumers, who instead preferred reactive actions, such as mitigation and remediation, over proactive identification of issues. For the Consumer Support category, consumers were significantly less likely to have chosen prevention and significantly more likely to choose mitigation than for the other three categories, suggesting that this preference is heightened in situations requiring support from the credit provider. The survey did not find significant statistical differences across demographic sub-groups, such as education, occupation, and household income. This indicates that consumers' expectations of firms' responses to harm are relatively consistent across diverse backgrounds and economic profiles.

## 4.2 Qualitative results

The results are presented in relation to either compliance or culture, along with generated sub-themes, supporting references and participant numbers in brackets.

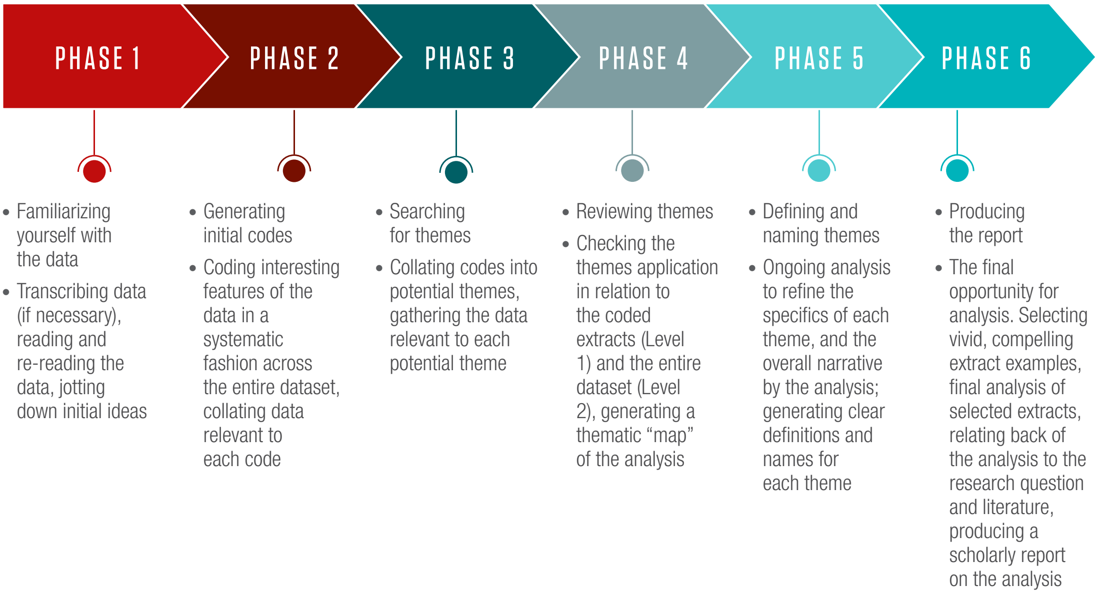
#### 4.2.1 COMPLIANCE BEHAVIORS

- **Interpretation:** practitioners stated that the first action they took was to interpret what the regulation expected of them and how they would be supervised. They described how interpreting the regulation went beyond the guidance, given that there was the need to consider how it applied to their firm: "The first is that it's still regulation. So it's written in regulatory language. So normal business people might struggle to understand it particularly well. You kind of need the compliance interpretation of it." [P6]
- **Education:** colleagues across the bank need to be educated on what needs to happen and why it needs to happen. There was an acknowledgement that banks can educate themselves by conducting customer research, especially in the context of vulnerable customers: "a lot about what we do is … around convincing yourself that what you do today is acceptable, and therefore not acknowledging the need to culturally … be doing something different in order to improve that customer experience, especially for those that are older, more vulnerable." [P1]
- **Use of data:** once data was collected, practitioners highlighted the process of converting it into insight and using it to drive change and evidence compliance: "When all is said and done, … how do we simply put, what we do and how does that align to the Duty and I think that… if you can't say it simply then you're obviously not doing it." [P2]
- **Prioritization and planning:** to prepare for meeting the regulation, practitioners stated that the allocation of dedicated resources and budget was essential. The importance of planning was also mentioned, due to the limited time scales: "I and my team were hired into the Consumer Duty roles as the first line of defense and the view was that the first line of defense will put together the business plan that's … all the actions and workflows that we need to complete by, we set ourselves a deadline at the end of March." [P6]

#### 4.2.2 CULTURE CHANGE BEHAVIORS

- **Collaboration:** practitioners highlighted the importance of collaboration, both across internal teams and externally with other firms and with the regulator: "I think that there's opportunity, where there is no competition or market risk … My point being that either through firms or trade bodies, such as UK finance, for example. And to the FCA, there's opportunities with things like this, to define collectively at industry level what good looks like." [P1]
- **Update internal processes:** for culture change to occur, banks need to invest in, and prioritize culture change, which can be done through utilizing technology, data, and communication with customers. It is also important to adopt an iterative approach to new processes, learning from experience, and evidence change to highlight benefits: "To fully implement Consumer Duty …, you know, it's not a one and done. It's … very much an evolving process." [P4]
- **Change in mindset:** acknowledging the need for change and adopting a non-economic viewpoint/mindset compared to one that traditionally values returns over outcomes for customers was highlighted by practitioners. For this to occur, it is important to create psychological safety in teams, whereby colleagues feel comfortable speaking up, challenging each other, and producing new ideas: "So I think there was definitely like, a lot of talk around… not having that culture where people feel comfortable to speak up or people… have that customer-centric mindset." [P9]
- **Embedding:** practitioners acknowledged that the whole firm must be mobilized to change, and a long-term culture shift requires all teams to embed the Consumer Duty into every piece of work. There is also a need to align the change with the bank's long-term strategy: "But the whole point of Consumer Duty is it's embedded at every level of the organization." [P6]

**Figure 5:** Braun and Clarke's (2006) six stages of thematic analysis




### 4.2.3 SUMMARY

Finance professionals drew a clear distinction between actions required for compliance with the Consumer Duty by the deadline and those necessary to instigate a long-term, firm-wide cultural shift. Regarding compliance, their efforts were primarily focused on education, acknowledging that banks could enhance their knowledge through participation in training sessions, and conducting primary or secondary research. Proactive planning, prioritization, and data-driven insights were also emphasized as key factors for compliance. Despite the FCA's outcomes-based approach to the Consumer Duty, compliance behaviors appeared process-driven, centered around identifying the necessary actions and providing evidence of their completion. For culture change, professionals emphasized the significance of collaboration with colleagues, other firms, and regulatory bodies to collectively define success. Compared to consumers' preferences, the behaviors of professionals exhibited more breadth across the firm and had an equal impact on all customers. This encompassed adopting a long-term change in mindset and embedding the principles of the Duty into every aspect of their work. Few participants mentioned seeking consumer input on desired firm behaviors, and it did not emerge as a prominent theme for either compliance or culture change. Consequently, the target behavior of conducting consumer research was chosen to bridge the gap between consumers and professionals and promote positive customer outcomes.

## 5. RESEARCH QUESTION 2: WHAT ARE THE BARRIERS AND FACILITATORS TO THE TARGET BEHAVIOR, AND HOW CAN THESE BE MODIFIED THROUGH "INTERVENTION FUNCTIONS" AND "BEHAVIOR CHANGE TECHNIQUES"?

To identify barriers and facilitators of the target behavior, the interview data were analyzed using Fereday and Muir-Cochrane's (2006) deductive thematic analysis method based on the COM-B model (Michie et al., 2014) and Braun and Clarke's (2006) six-phase thematic analysis process. As the interview questions regarded general challenges to implementation, responses were only coded if they related to the target behavior of conducting research with consumers.

To apply the COM-B model to the data, a code manual was created whereby each element of the COM-B model had a label, a definition, and a description of qualifications and exclusions to demonstrate when the code occurs [Boyatzis (1998)]. The transcripts were then coded deductively using the code manual. To ensure reliability, the first interview was analyzed by trustees at the Fairbanking Foundation. Any discrepancies were discussed, a conclusion was drawn which satisfied both coders, and the code manual was modified to reflect the change before continuing to code the other interviews.

Following deductive coding, data were then coded inductively to identify relevant themes under each COM-B component. Initial codes were generated by labeling each smallest possible data fragment with a descriptive unit of meaning, this time in relation to a barrier or enabler to the target behavior. Initially, 571 codes were identified, but after three iterative rounds of restructuring, similar codes were consolidated, and 172 codes remained [Miles et al. (2014)]. Codes were then collated into themes that could be categorized within each COM-B domain, which became the overarching theme.

Themes were reviewed, second-coded, and refined as described in the first round of coding. Intervention functions and behavior change techniques were identified based on the relevant COM-B influences, and these were analyzed using the APEASE criteria to suggest the most suitable techniques for influencing practitioners' behavior.

## 5.1 Results

The results are presented in relation to each COM-B component, along with generated sub-themes, supporting references and participant numbers in brackets.

## 5.2 Capability

### 5.2.1 PSYCHOLOGICAL CAPABILITY

Practitioners frequently mentioned the difficulties they faced in interpreting the regulation and understanding what was required of them by the FCA: "But we don't actually know and there's nothing in the regs that tells us specifically, we can't do this or we should do this." This was due to the guidance being ambiguous and unclear, with participants recognizing that the FCA's thinking was developing alongside theirs. Participants suggested that the guidance lacked examples of best practices that would have provided tangible insight into what they should be doing: "It's certainly weak in many aspects. It's a little unclear, I guess what good looks like." [P6; P9]

Even once the bank had collected data or research, there was a lack of understanding of how to use it: "What is difficult is to turn that into insight. So the key things that really give you confidence around the outcome you're delivering." Lack of knowledge and skills was a barrier for colleagues at all levels, driven by the absence of learning from experience and a traditional mindset that prioritizes returns over all else: "So why should we invest lots of money to do Consumer Duty as best as we possibly can if ultimately it's going to drag down returns?" [P3; P7]

## 5.3 Opportunity

### 5.3.1 PHYSICAL OPPORTUNITY

Physical opportunity barriers included limited budget, resources, and time, meaning that other activities are prioritized over research with customers: "Now, the latter i.e. the data, that requires material investment, okay, and I'll stop short of saying how many materials but material investment okay,". This was heightened by the unexpected magnitude of the regulation: "Consumer Duty has been really challenging because well, firstly, the breadth of it, the difficulty in defining scope, but the need for like strategic decisions on everything." [P1; P5]

Many practitioners stated they were working on the Consumer Duty alongside their usual responsibilities: "It was long days, it was you know, you need to kind of do your BAU stuff on top of doing this project on the side." The time limitations lead practitioners to prioritize the minimum requirements for compliance by July over wider changes, such as conducting consumer research: "But I think right now in a way, I think we're just doing work so that we meet that July deadline. And then everything else will probably be in the future." [P10; P9]

### 5.3.2 SOCIAL OPPORTUNITY

Working with colleagues and receiving support from senior managers were key factors in influencing practitioners' behavior. Participants mentioned a lack of accountability, and a need to share accountability across the bank and work collaboratively: "We try to ensure that we have really good relationships and we've tried to work towards them and sort of negotiate and compromise on something if they disagree with us." [P9]

A lack of support from senior managers was cited as a key barrier, linked to the capability of the managers in their understanding and prioritization of what needs to change: "I want to know nothing at all unless there's a problem that you need my help solving." One participant mentioned issues with bureaucracy as a barrier to implementing change: "In those types of firms, especially in kind of big organizations, there is a lot of bureaucracy and red tape and probably multiple levels of approval before you can get something like that done." [P6; P10]

## 5.4 Motivation

### 5.4.1 REFLECTIVE MOTIVATION

Many participants mentioned that the lack of motivation to conduct customer research was related to the belief that there would not be much value gained in talking to customers: “So I don’t think sometimes asking a customer … what would a good outcome look like for you? I don’t think we’d necessarily get a great deal more learning than we’ve already got.” More generally, there was a lack of belief in the impact of the Duty, as many participants believed they were already compliant and, hence, there was no need to make any changes: “I think that comes back to that … misunderstanding and it almost being seen as a bit of an over and above, well we already do a lot of this why are we having to do it: another hoop to jump through from the regulator.” [P3; P11]

This was driven by regulatory fatigue due to the number of regulations to follow, and a lack of understanding of the motivation of the FCA for creating the Consumer Duty: “Why they’re dressing this up as kind of a new duty is a little bit yeah, the cynic in me just thinks it’s, you know, we need to justify our fees and there is definitely a political angle to that.” [P10]

### 5.4.2 AUTOMATIC MOTIVATION

Unclear expectations of enforcement from the FCA led to a lack of automatic motivation from participants. Some participants expected the FCA to be strictly enforcing the Duty: “So we’re on notice that every initiative, every policy change, which might make things harder for customers, needs to be scrutinized through the lens of Consumer Duty. And if we don’t scrutinize it then the FCA will,” leading to a prioritization of compliance over culture change. [P3]

Other participants believed the FCA would not be enforcing the Duty: “There’s still this well the FCA is not going to come knocking on my door mentality because they won’t in all likelihood,” leading to a lack of motivation to change at all. [P6]

Most participants were unclear about how they would be supervised, while some mentioned that they expected it to differ by firm: “Some firms will gold plate what they already do and spend oodles and oodles and do things brilliantly and amazingly. Firms will do a little bit and then there’ll be other firms that do bugger all, let’s be honest.” This ambiguity in enforcement is driven by the lack of transparency from the FCA, decreasing firm’s trust in the regulator and demotivating them to comply: “I think where the lack of information or transparency from the FCA stems from now is we don’t really know how the model of supervision will change post-Consumer Duty.” [P1; P7]

## 5.5 Proposed “intervention functions” and “behavior change techniques”

To identify the most effective interventions to increase the target behavior and have the greatest effect on the outcome of good customer outcomes, the 20 barriers identified in the behavioral diagnosis were first assessed for need for change to ensure that targeting this barrier will have an impact.

This resulted in 11 influences that were mapped to intervention functions using the BCW matrix [Michie et al. (2014)]. Proposed intervention functions were evaluated using the APEASE criteria (affordability, practicability, effectiveness and cost-effectiveness, safety and side effects, and equity), to select the most suitable, and these were operationalized using frequently used behavior change techniques.

## 5.6 Summary

The results identified psychological capability as the COM-B component with the most barriers to conducting consumer research. Participants mentioned they had trouble understanding and interpreting the regulation, and so did not see how consumer research would be valuable. Even if they did identify this as an action, they were not aware of how to effectively apply any knowledge gained. This aligns with instrumental theories of compliance, as professionals cannot conduct a cost-benefit analysis without full knowledge of the outcomes of the behavior [Becker (1968)]. Similarly, Zeng and Botella-Carrubi (2023) found that practitioners usually consider consumer research as being “irrelevant” in guiding their decision making, since they do not have the knowledge or skills to understand the impact of the research.

Rousseau (2006) suggested that "Big E" evidence – generalizable knowledge of research methodologies – combined with "little e" evidence – knowledge of a particular context (in this case, consumers of finance products) – is required for successful research. In this study, the BCTs of "information about social and environmental consequences", "instruction on how to perform the behavior", and "action planning" were suggested. From the top, firms can educate their employees on how consumer research is a valuable activity in the context of the regulation and provide training on how to get the most from consumer research.

Another psychological capability barrier was a lack of learning from experience, despite the regulatory change not being novel or unexpected. Cannon and Edmondson (2005) suggest that complex organizations have difficulties learning from failure. This is due to systemic barriers, such as poor detection of failures and a lack of skills to extract lessons, and social barriers, such as the tendency for organizations to penalize failures and lack of skills for discussion and analysis. In this study, the BCTs of "instruction on how to perform the behavior" and "review outcome goal(s)" were suggested as firms should provide training on soft skills, including the ability to learn and adapt, and teams can reinforce learnings through review sessions at the end of each project and planning sessions at the start. The final psychological capability barrier was a lack of prioritization. Planning is important in translating intentions into behavior, as it encourages individuals to think about what they need to do in order to change [Sniehotta (2009)]. This study suggested the BCTs of "action planning" and "prompt/cues" to target this barrier. Teams should also include consumer research as an element of the implementation plan, including context, frequency, and duration, as well as implementing reminders to schedule research to ensure it is prioritized.

Lack of timely clarification and guidance from the FCA emerged as physical opportunity barriers to the target behavior. The more precisely behaviors are specified, the more likely they are to be conducted [Michie and Johnston (2004)], and the behavior is not specifically mentioned in the Consumer Duty guidelines. Professionals mentioned that they were attempting to reach out for clarification but were struggling to receive this. The BCTs of "restructuring the social environment" and "feedback on outcome(s) of behavior" were suggested. Audit and feedback is a strategy used in healthcare to change practice, whereby practitioners are shown how they are currently performing compared to explicit criteria or standards. Actions are then identified to establish how to improve performance [Jamtvedt et al. (2019)]. In the context of healthcare regulatory change, one meta-analysis found that on average, audit and feedback produced a median of 4.3% improvement in compliance [Ivers et al. (2012)], and financial regulations such as the Consumer Duty provide opportunities for these techniques to be applied in other contexts.

Social opportunity barriers such as lack of support from senior managers, lack of collaboration with colleagues, and lack of shared accountability align with normative theories of compliance, whereby decisions are influenced by beliefs, values, and norms that stem from social identity [Reus-Smit (2011)]. Zeng and Botella-Carrubi (2023) found a lack of engagement from stakeholders to be a barrier to conducting consumer research in practice, adding that team members often do not recognize their roles and responsibilities. They further elaborated that this could be a result of diversity in the team members' academic backgrounds, cultures, and disciplines. This study suggests the BCT of "restructuring the social environment" to enable employees to spend time with colleagues when prioritizing and planning, to facilitate support for, and collaboration, on consumer research. "Reframing" can assist with this, as advocates can convince others by drawing attention to the benefits of consumer research rather than the financial or time restraints. "Prompts/cues" and "demonstration of the behavior" also assist as teams can schedule regular check-ins or knowledge shares to facilitate collaboration.

Motivational barriers to conducting consumer research include the belief that the behavior lacks value; a resistance to change; and a lack of trust in the regulator. These barriers imply that beliefs and emotions are drivers of behavior and provide support for the application of behavioral science to understanding compliance. Some findings support the role of beliefs on compliance behaviors. Wenzel (2017) found that individuals consider their perception of fairness when deciding whether to comply, and in the context of tax evasion Enachescu et al. (2019) found that emotional experiences play a role in decisions; however, little prior research has been conducted on regulatory compliance of professionals where the outcome has little direct personal impact. To address reflective motivational barriers, this study suggests the BCTs of "information about social and environmental consequences" and "credible source" as firms can educate their employees on how consumer research is a valuable activity in the context of the regulation, which should come from a credible and respected source within the bank; and "social comparison" as colleagues can draw attention to the value brought by others who have conducted consumer research. To target automatic motivation, the FCA can increase trust by "restructuring the social environment" so that it is easier for firms to get in touch and using "prompts/cues" to schedule regular information sharing.

## 6. CONCLUSION

This study has effectively applied behavioral science frameworks to compliance with financial regulation, aiming to improve outcomes for customers under the FCA's Consumer Duty. Surveys with consumers and interviews with practitioners led to the conclusion that understanding consumers' views and priorities should be an essential feature of firms' implementation plans and that consumer insight should be used to drive decision making in response to identified poor outcomes. The suggested intervention functions and BCTs offer valuable insights for firms to enhance their compliance efforts and align consumer and professional priorities effectively. Applying these findings in practice and conducting further research to address the identified limitations will play a crucial role in fostering a consumer-centric culture within financial institutions and achieving positive outcomes for customers.

## REFERENCES


Becker, G. S., 1968, Crime and punishment: an economic approach," Journal of Political Economy 76:2, 169-217

Boyatzis, R. E., 1998, Transforming qualitative information: thematic analysis and code development, SAGE

Braun, V., and V. Clarke, 2006, "Using thematic analysis in psychology," Qualitative Research in Psychology 3:2, 77–101

Cannon, M. D., and A. C. Edmondson, 2005, "Failing to learn and learning to fail (intelligently)," Long Range Planning 38:3, 299–319

Enachescu, J., J. Olsen, C. Kogler, M. Zeelenberg, S. M. Breugelmans, and F. Kirchler, 2019, "The role of emotions in tax compliance behavior: a mixed-methods approach," Journal of Economic Psychology 74, 102194

Fereday, J., and E. Muir-Cochrane, 2006, "Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development," International Journal of Qualitative Methods 5:1, 80-92

FCA, 2022a, "PS22/9: a new consumer duty," Financial Conduct Authority, https://tinyurl.com/mf2yjasp

FCA, 2022b, "Strategic review of retail banking business models," Financial Conduct Authority, https://tinyurl.com/my9zrm86

Handley, M. A., A. Gorukanti, and A. Cattamanchi, 2016, "Strategies for implementing implementation science: a methodological overview," Emergency Medicine Journal 33:9, 660-664

Hudson, B., D. J. Hunter, and S. Peckham, 2019, "Policy failure and the policy-implementation gap: can policy support programs help?" Policy Design and Practice 2:1, 1-14

Ivers, N., G. Jamtvedt, S. Flottorp, J. M. Young, J. Odgaard-Jensen, S. D. French, M. O'Brien, M. Johansen, J. M. Grimshaw, and A. D. Oxman, 2012, "Audit and feedback: effects on professional practice and healthcare outcomes," Cochrane Database of Systematic Reviews, https://tinyurl.com/2p8b9sh3

Jamtvedt, G., S. Flottorp, and N. Ivers, 2019, "Audit and feedback as a quality strategy," in Busse, R., N. Klazinga, D. Panteli, and W. Quentin (eds.), Improving healthcare quality in Europe: characteristics, effectiveness and implementation of different strategies, World Health Organisation

Michie, S., and M. Johnston, 2004, "Changing clinical behaviour by making guidelines specific," British Medical Journal 328:7435, 343-345

Michie, S., L. Atkins, and R. West, 2014, The behaviour change wheel: a guide to designing interventions, Silverback Publishing

Miles, M. B., A. M. Huberman, and J. Saldana, 2014, Qualitative data analysis, SAGE

Peat, D. M., 2021, "Perception and process: towards a behavioural theory of compliance," Journal of International Dispute Settlement 13:2, 179–209

Reus-Smit, C., 2011, "Obligation through practice," International Theory 3:2, 339-347

Rousseau, D. M., 2006, "Is there such a thing as "Evidence-Based Management"?" Academy of Management Review 31:2, 256-269

Sniehotta, F. F., 2009, "Towards a theory of intentional behaviour change: plans, planning, and self-regulation," British Journal of Health Psychology 14:2, 261-273

Wenzel, M., 2017, Tax compliance and the psychology of justice: mapping the field, Routledge Books

WBF, 2023, Preparing for the consumer duty – priorities for implementation and compliance, Westminster Business Forum Policy Conference: Next Steps for the Consumer Duty

Zeng, T., and D. Botella-Carrubi, 2023, "Improving societal benefit through transformative consumer research: a descriptive review," Technological Forecasting and Social Change 190, 122435

# GENERATIVE ARTIFICIAL INTELLIGENCE ASSESSED FOR ASSET MANAGEMENT[1]

**UDO MILKAU** | Digital Counsellor


## ABSTRACT

Since the launch of the generative artificial intelligence tool ChatGPT end of 2022, there has been an incredible public awareness. Doomers predicted an end to humanity, while more reasonable assessments discussed the impact on traditional industries and on the workforce. In a nutshell, generative artificial intelligence is nothing more than statistical estimation and continuation of an input sequence based on a text corpus of the past. To evaluate the actual impact of generative artificial intelligence (AI) and large language models (LLM), this paper uses the case of asset management as a benchmark. These statistical estimators can produce a "next best token" based on an "internet average", i.e., tremendous text corpora gathered from internet sources, but which cannot understand, predict anything new, or create something innovative. Consequently, generative AI/LMMs can augment staff to perform "on average", or help internet users obtain an "average answer" to their questions about financial management. While this can (and probably will) change the future landscape of financial advice and the way consumers access information, generative AIs/LMMs are far from any type of "superintelligence". The potential of misuse by human actors, however, remains the biggest danger and has to be monitored closely.


## 1. INTRODUCTION: SHOULD WE FEAR GENERATIVE ARTIFICIAL INTELLIGENCE?

Generative artificial intelligence (AI) is a topic of much debate. Between March and May 2023, three open letters [Clarkson Law Firm (2023)] have been published about the existential risk to humanity posed by generative AIs. Bengio et al. (2023) ask: "Should we risk loss of control of our civilization?" Altman et al. (2023) state that "superintelligence will be more powerful than other technologies humanity has had to contend with in the past. … Given the possibility of existential risk, we can't just be reactive. Nuclear energy is a commonly used historical example." And Hinton et al. (2023) suggest that "mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."

Warnings that "AI-based agents could achieve their individual goals beyond any human control" are not new – from D. F. Jones' 1966 science fiction novel "Colossus" to the "Terminator". However, it should be noted that all computer codes written by humans are – at least until today – fully deterministic and merely executing pre-defined "if – then – else" statements. Any contemporary AI is only "able to fit a function to a collection of historical data points" [Pearl and Mackenzie (2018)], which might be highly complicated math but is nothing more than zeros and ones on a Turing engine.

This is not to say we should be complacent. As Shevlane et al. (2023) state: "As AI progress has advanced, general-purpose AI systems have tended to display new and hard-to-forecast capabilities ... Future systems may display even more dangerous emergent capabilities, such as the ability to conduct offensive cyber operations, manipulate people through conversation, or provide actionable instructions on conducting

[1] The status of the development of AI and generative AI analyzed in this article is as of August 2023. The financial products mentioned are not meant as financial advice but as illustrative examples only.

acts of terrorism." It is possible that chatbots like OpenAI's ChatGPT, Google's Bard [Krawczyk and Subramanya (2023)], Meta's LLAMA 2-Chat [Touvron et al. (2023b)], Anthropic's Claude 2 [Anthropic (2023)], or Aleph Alpha's (2023) Lumi/Luminous could be misused for disinformation.

Despite such risks, few can deny the potential benefits that generative AI could unleash. In this article, we intend to look at how, and whether, it can impact the asset management industry.

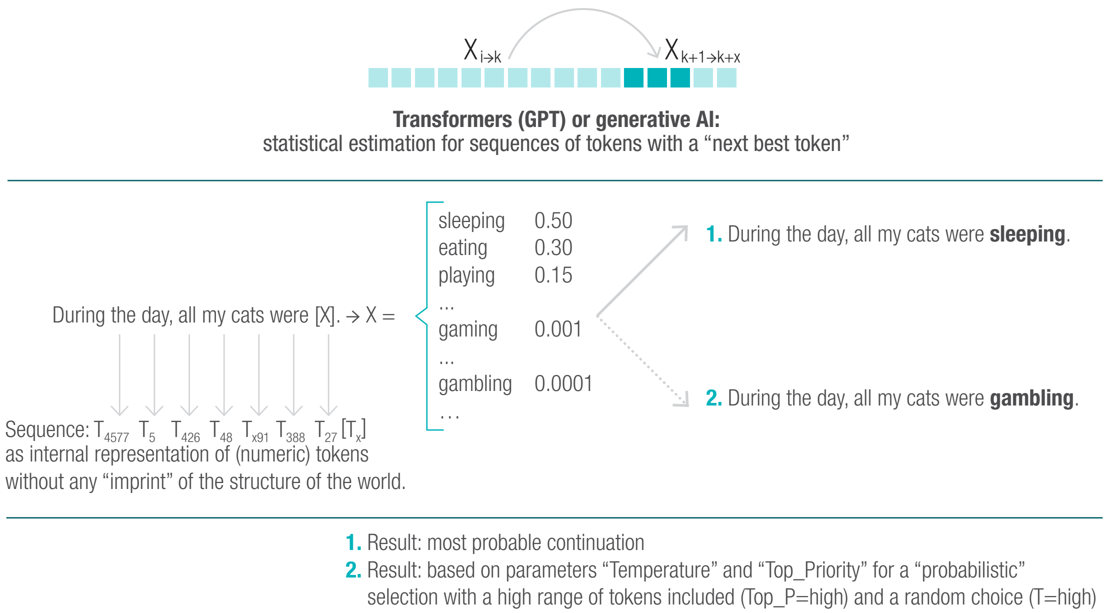
## 2. GENERATIVE AI IN A NUTSHELL

A schematical description of generative AI as a tool for the continuation of sequences of tokens with statistical probabilities is given in Figure 1. This description avoids the mathematical details but illustrates the basics: generative AI estimates a "next best token" based on the ex-ante fitted statistical probabilities of a tremendous number of existing sequences as inputs for "training" of internal parameters (called "learning" in the mathematical sense of the so-called "statistical learners") [Shalev-Shwartz and Ben-David (2014)].

The sequences can be any series of tokens: from words in a text (depending on the language potentially with pre-suffixes as additional tokens) or computer code to mathematical calculations[2] or the sequence of moves of a robotic arm. The most typical implementation of generative AI is "text-to-text"[3] and, consequently the term "large language model" (LMM) is used synonymously. Current systems like OpenAI's GPT-4, Meta AI's LLAMA 2, or Google AI's PaLM2, incorporate hundreds of billion parameters, are "trained" with trillions of tokens, and include so-called "multimodal" inputs[4] with dedicated text about computer codes, mathematical tables, or text-based games [OpenAI (2023), Touvron et al. (2023b), Ghahramani (2023), Amazon (2023)].

Despite the additional "fine-tuning" with mathematical calculations, Choi (2023) suggests that GPT-4 is not able to "solve" a simple multiplication when prompted to multiply 999 by 876. Although LLMs show improved capabilities when scaled to some hundred billion parameters and trillions of tokens, pure LLMs remain statistical estimators of the "next best token". They can be complemented in a hybrid way with rule-based systems, where one could attach a "pocket calculator" to do the math correctly, but there are practical limitations to few add-ons.

**Figure 1:** Generative AI as a system for continuation of sequences of tokens with statistical probabilities




The system represents statistical correlations between tokens in the corpus used for "training", but no "imprint" of the structure of the world. Two results for different settings of control parameters are shown, which illustrate that the system does not represent any "meaning".

[2] Like '13' '+' '16' '=' ?; with each object 'x' being a token but without any mathematical understanding or any difference between a variable or a number.
[3] For simplicity, more complicated approaches for text-to-image/visual/video generation from text prompts (e.g., OpenAI's DALL-E), image processing (e.g., Adobe's Firefly) will not be covered in this article.
[4] A recent research paper by Microsoft Research [Bubeck et al. (2023)] claims to see "Sparks of artificial general intelligence' in GPT-4."

During the ex-ante step of parametrization ("training" or "learning"), the input data will be used to fit a high-dimensional probability distribution to all sequences of tokens. Only for some systems [e.g., LLAMA-1; Touvron et al. (2023a)] the sources – like archives from web crawlers, Wikipedia, GitHub, books, etc. – are known, but not for GPT-4/ChatGPT or LLAMA 2/LLAMA 2-Chat. LLMs are statistical representations of the input text corpus,[5] hence contains all of the errors, biases, and disinformation of the input. After a model is "trained", it is a static implementation – without any "self-learning", "adaptation", or "in-context-learning".[6]

At run-time, the user can "prompt" the model with a new sequence (e.g., a sentence with a missing word or an equivalent question asking for this word), which should be completed with the "next best token" (i.e., the missing word or answer). In Figure 1, the sequence to be completed is: "During the day, all my cats were [x]", where [x] denotes the missing token. Internally, the model does not process words, but representations of tokens such as $T_{4577}$ $T_5$ $T_{426}$, $T_{48}$ $T_{x91}$ $T_{388}$ $T_{27}$ [$T_x$]. In this simplified description, the model would contain a list of words (or tokens) with statistical probabilities to fit in the [x]: sleeping, eating, playing and so on. The output could either be a result with the most probable continuation or dependent on the control parameters "Temperature" and "Top_Priority" for a selection with a high range of tokens included (Top_P=high) and a random choice (T=high). Such a choice by the provider[7] of the model generates more "creative" results for repeated "prompts", but could end up with "During the day, all my cats were gambling" – making the result similar to Lewis Carroll's "Alice's Adventures in Wonderland". It is essential to understand that LLMs are based on probabilities, and sometimes on errors if included in the input dataset, and that the same prompt could generate different answers, like in the cat-example above, depending on the setting of the control parameters (and also if the post-processing is "re-tuned" by the providers from time to time).

So-called chatbots like OpenAI's ChatGPT, Google's Bard, or Meta's LLAMA 2-Chat are optimized front-ends for users' interaction. It is quite impressive how much text can be generated with the rather simple concept of continuation of input sequences: from summaries of input text or text search to whole essays about a given subject. However, with trillions of "trained" parameters, LMMs are sophisticated statistical representations of the text corpus.

There are serious discussions about whether ChatGPT (currently less discussed for other generative AI tools) should be allowed in schools,[8] colleges, and universities. Because the average user cannot modify the setting of control parameters – or even is not aware of these parameters – such tools reveal limitations. A student wants a problem in programming to be solved "correctly", an essay written with accurate facts but some "personal" wording, and a poem created with "creativity".

While different styles for – typically repetitive – homework can be achieved by so-called "prompt engineering" (adding more "direction" to the question with keywords such as "act as"[9] and/or trying some iterations of prompts), the basic capability is always a statistical LLM for a "next best token", without any understanding of content and context.

Due to the nature of LLMs as "statistical estimators", the parametrization follows the Central Limit Theorem, as the input consists of statistically independent sequences and the result is – simplified – the mean probability[10]. Subsequently, all common errors, wide-spread disinformation, and shared nonsense obtain high probabilities to be included in the output. As long as LLMs represent text corpus from the internet, they follow the "Dr. House Theorem": "It's a basic truth of the human condition that everybody lies. The only variable is about what" [Dr. House (2005)]. Furthermore, LLMs cannot detect any "errors" internally, and corrections have to be made in dedicated post-processing steps (especially with so-called Reinforcement Learning from Human Feedback "RLHF").

[5] Grossmann et al. (2023) suggest that it might make sense to use LLMs in social science research instead of asking or testing real persons. Bai. et al. (2022) proposed a "fine-tuning" of LLMs according to the beliefs of human supervisors of social science studies.
[6] The term "in-context learning" is a terminus technicus, which has nothing to do with true learning.
[7] The public source code of Meta's LLAMA 2-Chat provides direct access to Top_P and Temperature parameters via the dedicated API.
[8] In the technical paper about GPT-4 (OpenAI, 2023), GPT-4 is benchmarked inter alia with so-called "AP exams" (Advanced Placement exams, in which students can earn college credit while still in high school). GPT-4 performs with high scores, in the above 80th percentile, especially for tests focused on writing summaries like in history or art.
[9] Prompt engineering with "act as …" is sometimes referred as "commands" or as "creating personas". An example could be: "**Act as** a financial advisor with a focus on long-term investments ..." before prompting "Construct a balanced portfolio of ten securities for a risk-averse investor of age 40!" Nevertheless, these commands are not programming language commands, but improved prompts. Similar approaches are known as "chain of thought prompting", especially when mathematical questions are parsed into elementary steps or simple algebra.
[10] For situations with only a few, or even singular, references in the text corpus, the result will be a somewhat modified retrieval of these original sources, or even plagiarism.

Current developments like Google AI's PaLM2 [Ghahramani (2023)] have been extended with more content – especially more domain-specific input, such as multilingual samples (up to the ability to translate idioms or jokes), scientific papers, mathematical expressions, or pre-training on publicly available source code in programming languages from Python and JavaScript to Prolog and Fortran. With these implicit "rules", such as lists of mathematical calculations, formulas, or programming templates, such extended LLMs are crossing the frontier to former "expert systems" as part of symbolic-logic approaches of AI of the 1950s to 1980s. Similar fine-tuning has been made to generate automated commentary with statistics and language of sports [IBM (2023)].

## 3. WORK PERFORMANCE AND AUGMENTATION OF LOW-SKILLED STAFF

With asset management as a litmus test, generative AI-chatbots were accessed with the prompt: "Construct a balanced portfolio of ten securities for a risk-averse investor of age 40". Such prompts are sometimes described as "zero-shot prompting",[11] as no additional "help" was given to ChatGPT, Bard, and Llama 2-Chatbot. For specific tasks, it can be necessary to provide more "direction" in the prompt, such as a list with Q&A plus one open question to guide the generative AI towards the specific subject or to iterate prompts with supplementary information.[12]

The result shown in Table A1 in the Appendix is astonishing. Without any "understanding" of asset management, ChatGPT provides a result, which is convincing at first glance. One could discuss whether a simple portfolio of index funds and Exchange Traded Funds (ETFs) would be a "correct answer", but it is not unrealistic at all. However, I made two basic checks. Firstly, I checked to see whether the funds and ETFs recommended actually exist. Secondly, given that ChatGPT is limited to input until 2021 – what would a Google-search for "best index funds" or "best ETFs" in June 2023 offer as result, if two references are selected rather randomly from the list of search results?

Google's Bard – a combination of an LMM and a code-based system including direct access to online search – provides a similar portfolio of index funds and ETFs, but only those provided by a single company. In comparison, Meta's Llama 2-Chatbot (accessed via llama2.ai few days after the public release) is much more "restrained", providing only a textbook-like structure of a balanced portfolio, and explicitly referred to "consult with a financial advisor".

The synopsis for ChatGPT and Bard is shown in Table A2 in the Appendix. We find that all the ETFs and index funds mentioned by ChatGPT exist. The result of the ChatGPT is similar to, and in some cases overlapping with, some rankings found via Google search (i.e., Forbes and Yahoo). While Bard also provides a number of overlapping results (with ChatGPT and with Forbes and Yahoo), it only offers Vanguard index funds and ETFs. It is hard to know why it only chooses this specific company. Meta's Llama 2-Chatbot differs in many ways: it is an open-source software (accessed via an implementation at llama2.ai), the parameters "Temperature" and "Top_P" can be adjusted by the user (the default was used with T= 0.10 and Top_P = 0.90), and the results are rather textbook-like, while ignoring the requirement in the prompt to use "ten securities". It would require more work to check for different settings of the control parameters, but the default setting provided a rather generic result without taking the risk of providing a specific answer.

The three examples (generated during June/July 2023) reveal that in the default setting the generative AI tools produce rather different results for the same prompt. Personally, I would rank the results provided by ChatGPT as being on par with an average finance journalist. Bard, on the other hand, should be more up-to-date, though for some unknown reason shows a strange bias to one single provider. However, given that Bard is still in development this bias might be overcome in future versions. Nonetheless, these examples show that the tremendous text corpus[13] derived from the internet used as input to ChatGPT contains much input on a specific subject matter, such as asset management, that enables "generative" text production based on statistical correlations only.

[11] The terminus technicus "zero-shut" is somehow misleading, as "zero-shot learning" is a special case for pattern recognition with deep learning, when the new event belongs to a class of events not "learned" (and, therefore, not classified in the model) but can be interpolated between learned classes.

[12] For example, such "in-context learning" has been applied to the "continuation" of a prompt in English with a mathematical problem plus few examples of "translation" into a formal code (in "Isabelle" language), for which the LLMs are trained inter alia with libraries with Isabelle code.

[13] The idea that scaling AI models would improve quality is taken as an axiom. Nonetheless, Gigerenzer (2023) contests this approach and argues that smaller, more traditional and "explainable" AI models could have a much better trade-off, especially in situations with high complexity.

While these statistical correlations within the text corpus are sometimes described as "emerging",[14] "dynamical", "self-learned, or "adaptive", a generative AI tool does not act "autonomously".[15] Currently, and with an average user as benchmark, ChatGPT (and to a lesser degree Bard or Llama 2-Chatbot) seems to "perform" similarly to many Robo-advisors,[16] but, of course, lacks the possibility of order execution or saving plans. There is a danger, however, that users would perceive such a "conversational technology" as empathic and emotional,[17] because it emulates the (statistical) features of human conversation. There is an additional danger that generative AI tools could be used maliciously to publish "deepfakes" or disinformation for the purpose of manipulation, misconduct, or fraud.

A recent online survey with ten thousand consumers in 13 industrial states conducted by Capgemini (2023) revealed that 51% of those online-affine respondents answered yes to the question: "I am aware of the latest trends in the generative AI space and have also explored tools such as ChatGPT, DALL-E". More than half of the respondents trust generative AI to "assist with financial planning", and two-thirds said that they could "benefit from receiving medical diagnoses and advice" from generative AI. While answers to online surveys might differ from actual behavior and only indicate a principal attitude, the rather positive acceptance of generative AI by online-savvy consumers indicates a significant potential for future generative AI-based financial advice[18,19]. It has to be made clear, however, whether this advice is given by a human advisor or a machine.

Consequently, it is appreciated that the European Parliament (2023b) proposed amendments to the European Commission's proposal on an Artificial Intelligence Act (AIA) with disclosure obligations for AI-generated content. Likewise, U.S. President Biden [White House (2023)] announced voluntary commitments by leading AI firms to include watermarks on generative AI-generated content. It should be said that such disclosures do have their opponents as well [Altman et al. (2023), Worldcoin (2023)].

Brynjolfsson et al. (2023), who published a study on the impact of "generative AI assistants" on labor productivity, found that customer support/call center agents that were augmented by generative AI experienced a 14% productivity increase on average, as measured by issues resolved per hour, though there was some spread according to the skill level. It should be added that a number of issues can be handled automatically before a human agent is needed. First level solutions from interactive phone systems to rule-based text-chatbots can solve a number of customer issues, such as password requests and changes, changing of address, and account balance. But for the second level, when human agents are needed, the study is consistent with two other findings that AI can augment lower performance within a certain job profile.

Kanazawa et al. (2022), who studied the impact of AI on worker productivity in the context of taxi drivers, found that an AI "assistant" reduced the time spent on cruising by 5.1% using the full sample, but with all the gains concentrated on low-skilled drivers, narrowing the productivity gap with high-skilled taxi drivers. And in a recent online experiment with preregistered college-educated professionals randomly exposed to ChatGPT, Noy and Zhang (2023) found that "the generative writing tool increased the output quality of low ability workers and reduced time spent on tasks for workers of all ability levels."

There are already some implementations with a focus on augmentation. The German federal state of Baden-Württemberg developed a text assistant "F13" for staff in public administration based on the Aleph Alpha's Luminous generative AI [StM.BW (2023)]. This first adaption of generative AI for public administration provides basic functionalities such as summaries of text inputs, except for confidential or personal data, generation of (short) notes from stored cabinet bills, and research in a knowledge base of information for public services. Given that Germany is predicted to have one million public administration vacancies by 2030, according to McKinsey, such basic text assistants can relieve staff from "mechanic" text writing, i.e., augmentation instead of substitution.

[14] One example is the recent work of Webb et al. (2023) about "emergent analogical reasoning", when generative AI is prompted to continue text-based sequences of numbers (aligned to so-called Raven's Standard Progressive Matrices). Nonetheless, this is still a continuation of a given sequence based on the statistics of the text corpus used as input (with similar examples described on many websites).

[15] Winograd and Flores (1986) published a seminal work about the fundamental differences between computer-based AI and human cognition.

[16] Typically, Robo-advisors provide a proposed portfolio of ETFs and/or index funds matching the customer's risk profile, personal experiences with capital markets, and individual expectations.

[17] This has to be separated from so-called "emotional AI", which claims to recognize the "big five" basic emotions with facial recognitions. The basic concept of a "facial action coding system" was developed by Ekman and Friesen (1978), which was based on many assumptions and ignored the fact that facial expressions are not independent from the socio-cultural background [Laajaj et al. (2019)].

[18] However, CFPB (2023) raised concerns about the fact that "Poorly deployed chatbots can impede customers from resolving problems."

[19] It is beyond the scope of this paper to discuss the issue of "social synchronization" when it comes to investment decisions, but there are links to problems of "social trading" [BaFin (2023)].

Concerning the quality of AI-based systems, Liu et al. (2019) compared "deep learning" (see Section 5) with healthcare professionals in detecting diseases from medical imaging. For medical tests[20] to predict diseases, they found a pooled sensitivity of 87.0% for deep learning models and 86.4% for healthcare professionals, and a pooled specificity of 92.5% for deep learning models and 90.5% for healthcare professionals. As deep learning for image recognition is based on the pooled "experience" of human professionals (i.e., datasets with images and diagnoses by humans as "labels"), it is plausible that it is emulating the quality of those professionals on average. However, detecting diseases from medical imaging is only one step in medical diagnosis and only one step of many for therapies. Deep learning tools can augment the diagnosis process, reduce the workload required to analyze the majority of images with "average" patterns, and can help healthcare staff to have more time for patients, but it cannot replace experts.

Concerning clinical applications of LLMs, Singhal et al. (2023) published results from an instruction-tuned variant of Google's PaLM called Flan-PaLM2 and an own instruction prompt tuned model "Med-PaLM" that attempted to answer questions from a multiple-choice dataset "MultiMedQA". The found that Flan-PaLM achieves a 67.6% accuracy on MedQA (U.S. Medical Licensing Exam-style questions), with Med-PaLM doing even better, but both were inferior to clinicians. In other words: the tested fine-tuned LMMs perform like students on multiple-choice questions, but not like the professionals.

## 4. STATISTICAL ESTIMATORS AND EFFICIENT MARKETS

Are there any "hidden" capabilities in generative AI and LLM that could be implemented to augment asset management or find an investment portfolio strategy? Could there be a way to either forecast stock prices in some kind of advanced chart technology or to find exceptions of performance (or of risk) in large-scale data collections? This use case, of course, comes with the assumption of reproducibility, i.e., without any setting of 'Temperature' and 'Top_P' for results at random and stable fine-tuning.

> *Generative AI tools make statistical estimations based on a continuation of "next best tokens" but without any chance to go "where no one has gone before".*

Generative AI/LLM neither work with market data nor real-time feeds but are parametrized on a "historic" internet text corpus. Based on the LLM capability to produce summaries for longer prompts based on statistical estimation of "next best token" (but not based on content or on context!), such tools can provide a summary of archived news feeds, a synopsis asked for by an input prompt, or an advanced search in analyst reports (if trained on this corpus of documents). However, the trend to the "statistical average" wipes out substantial information. In other words, generative AI cannot "generate" new information or novel insight but provides – simplified – statistical fits to the past.[21]

User-friendly generative AI front-ends offer the option to create marketing material, summaries, and presentations, but come with two downsides and one question. First, the input text corpus used for this specific tool is not disclosed (compared to other tools that provide sources) and, consequently, no checks on bias, copyright, etc. can be undertaken. Second, any prompt will be processed on the proprietary infrastructure, which could exploit vulnerabilities if confidential company data or protected personal data is used in the prompt. Third, any actual performance increase might be hard to detect, because one can find actual examples for which the effort to "engineer" the prompt take as long as the time saving by the tool. This problem requires more research like that undertaken by Brynjolfsson et al. (2023) to compare quantified performance increases[22] with overhyped marketing promises.

[20] For test results, the measures of prediction power (ex-ante) versus real classification of events (ex-post) are defined as: sensitivity = true positive/(true positive + false negative) and specificity = true negative/(true negative + false positive).

[21] A recent incident might illustrate the abilities and limitations of AI in general. Mid-August 2023, A Cruise autonomous vehicle (AV) had a collision with a fire truck in an emergency with its forward-facing red lights and siren on. According to Cruise (2023): "The AV positively identified the emergency vehicle almost immediately as it came into view ... The AV's ability to successfully chart the emergency vehicle's path was complicated by the fact that the emergency vehicle was in the oncoming lane of traffic, which it had moved into to bypass the red light." As AI-based systems are statistical classifiers, they are able to classify events according to "known" patterns derived from the training data but have limitations when "unexpected" situations occur.

[22] It would be beyond the scope of this article to review the application of generative AI to write computer code based on prompted descriptions of the problem, such as those included in tools like PaLM2 [Ghahramani (2023)], with a comparison of junior programmers augmented by generative AI versus senior experts and programmers augmented by generative AI versus software engineering tools such as so-called low-code programming.

For any attempt to "continue" a time series of prices, one has to take the "efficient market hypothesis" (EMH) into account.[23] Fama (2013), while discussing the "joint hypothesis problem", asks "Do the tests fail because the market is inefficient or because we have the wrong model for rational expected returns?" Tests of historical data reveal that there is no contradiction to the hypothesis that "efficient markets" directly process all (available) new information.[24] Louis Bachelier realized back in the 1900 that (efficient) markets always offer fair prices, because buyers and sellers agree to a transaction based on their individual evaluations of available information.

Despite that, a huge number of sources on the internet are publishing "predictions" on price developments of crypto coins and tokens, although they have no substantial foundations beside the "fear of missing out" (FOMO). Any generative AI/LLM using this text corpus as input could provide output to a prompt on "Please forecast the price development of Bitcoin!" with an average of the input statements. This is, of course, an extreme example, but it illustrates that generative AI/LLM trained with "scraped" data from the internet could initiate feedback-loops if decisions would be triggered by such "advice". One interesting development in this context is the solution ChatGPT™ provided by Ohio-based Futuri Media, LLC. RadioGPT™ is a three-step approach [Futuri (2023)], which applies a proprietary technology to scan social media like Facebook, Twitter, Instagram, etc., for topics with relevance to a local radio market, calls GPT-3 technology for text-to-script, and voice generation for script-to-audio. There is not much work left for journalists in this automated radio solution.

The results of Brynjolfsson et al. (2023) suggest that generative AI/LLM can augment (but not substitute) support staff but not provide additional value to experienced asset managers looking for real "alpha". When professional investors try to identify evolving patterns[25,26] information – from macroeconomic data and annual or quarterly reports to ad-hoc news and customers' evaluations in social media – can be analyzed automatically with algorithms, but this is beyond the capabilities of generative AI/LLM.

## 5. BEYOND GENERATIVE AI - BUT ALWAYS STATISTICAL CLASSIFIERS

Shalev-Shwartz and Ben-David (2014) wrote their book about understanding machine learning a mere nine years ago, however, it looks very different to the topics that are being discussed currently. Firstly, their book starts with a formal model about "the statistical learning framework", making clear that "machine learning" is a mathematical concept. Secondly, the algorithmic methods focused on "support vector machines" (SVM), kernel methods, decision trees and random forests, or nearest neighbors – with only twelve pages about artificial neural networks. Finally, the chapter about "generative models" started with an introduction that this book was mainly about a "discriminative approach", in which the goal is to "learn" (or parametrize) an accurate statistical predictor, whereas generative models aim to fit parameters to the underlying set of input data. The development of artificial neutral networks – with the basic "activation function" $\phi\,(\sum w_{ij} x_i)$ at every node and $\partial/\partial x$-optimized backpropagation to minimize a "loss function" alike linear regression [Milkau (2021)] – with tremendous amounts of input data needing immense computer resources changed the discussion.

The three main types of concurrent artificial neural networks (ANN) are summarized in Figure 2. An overview of the development until 2014 (before the development of the current generative AIs or "transformers") are given by Schmidhuber (2015) and LeCun et al. (2015).[27] This point of view, however, is primary technological and focused on the internal algorithms of the ANNs to obtain a statistical fit of parameters to the data.

An alternative perspective was presented by Johnson (2006), who states that "Computer systems and other artifacts have intentionality, the intentionality put into them by the intentional acts of their designers." This holds true for all existing AI, including generative AI, because they are deterministic computer codes, which are "trained" to fit a function to input data. They have neither intentionality, nor consciousness,

[23] It is worth noting that the EMH does not assume normal distributions of asset returns but allows so-called "stylized facts" as known statistical properties of empirical time series of observed asset returns (especially for short term correlation, etc.).

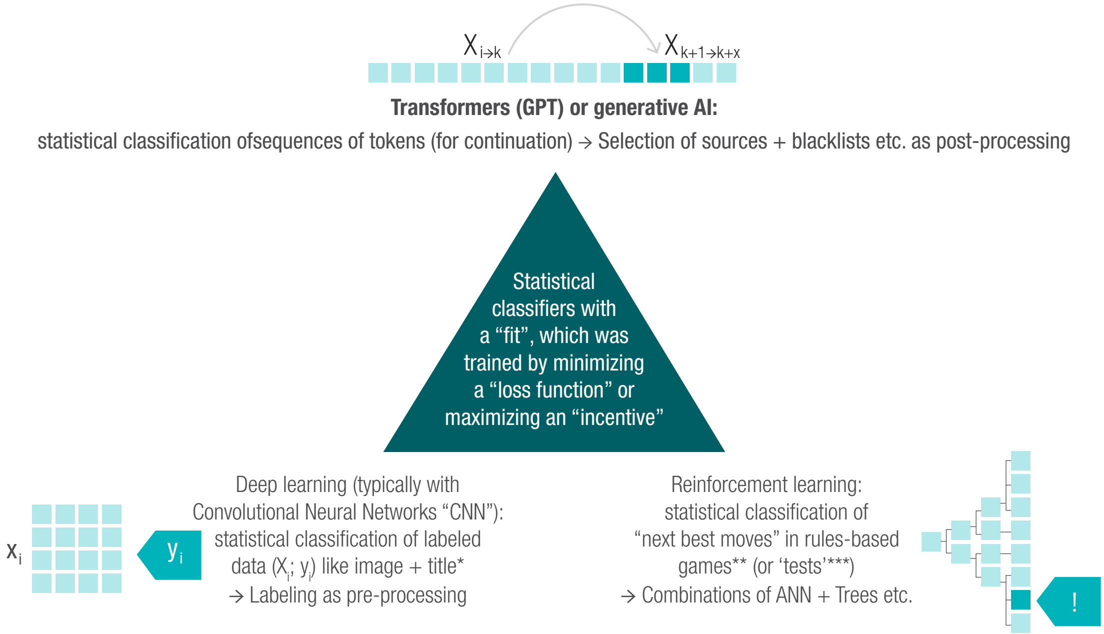
[24] It should be noted that this definition does not refer to any theoretical Pareto optimum of markets, but simply states that all available information is directly included in market prices.

[25] These tools can search for statistical outliers or to make statistical classifications between pre-defined classes of events. There is one fundamental problem that the number of recorded "negative" events is often dwarfed by the "positive" ones. While sufficient statistics about "negative" retail consumer credit events exist, this challenge can be obstacle for any statistical classification.

[26] A special topic is the so-called "sentiment analysis" from simple counting of ("positive" versus "negative") key words in company reports via news feeds to measurement of the time between questions and answers in investor conferences. The psychological-linguistic assumptions in these models can be questioned (including the question of what is "measured" in a statement made by an executive: the "public performance" of this executive or the financial performance of a company). On the other hand, sentiment analysis can be compared to chart analysis – always trying to find some extra beta in public information already included in the market evaluation.

[27] Deep learning or "reinforced learning" are also variants of ANNs. Reinforced learning was also applied in multi-agent models to simulate stock price formation [Lussange et al. (2022)] but not for prediction of future price developments.

**Figure 2:** The main types of "Artificial Intelligence" based on Artificial Neural Networks (ANN), which all belong to the universal category of "statistical classifiers"




Notes: *) In cases with a continuous labeling (numerical values) "zero-shot learning" is also possible for such classes, which can be interpolated. **) The optimization of parameters during the training will be based on the "score" reached at the end of a game as "incentive". ***) i.e., "recommendation engines" based on probabilistic tests of "next-best-offer" and monitoring of feedback typically do not require ANNs but simpler AI approaches.

autonomy, or "free will". Human programmers, on the other hand, implement their intention, define the scope especially by selection of the data for input, invest a lot of human effort for pre- and/or post-processing, and are responsible for the results [Milkau (2021)].

In case of deep learning for image/pattern recognition, one needs a huge dataset of images[28] plus the classifications (labels). While images can be taken from image databases or from image collections on the internet, they lack in many cases a (machine readable) label with the correct classifications. But this labeling requires a lot of human effort and limits the scope of the application. For example, pattern recognition "trained" with data of cats and dogs to classify between them (and in this extreme case not even with a third class "unknown", which would have been defined and trained ex-ante) will classify each new image as "cat" or "dog", although it might be a car – and it is perhaps not possible to explain[29] why one image of a car will be classified as cat and another image of a car as dog. Nonetheless, it is the responsibility of the human user of such a system to apply it as designed (i.e., for "cats" and "dogs" only).

In the case of reinforced learning,[30] typically two ANN systems are used in the first step of "training" to play a (rules-based) game against each other millions of times with an "incentive" for the winner. In a second step, the AI system is able to play this "repeated game" based on the parametrization, which was determined during the training but without any "self-learning" at run-time.[31] In many cases, these systems are a pre-configured configuration of ANNs plus tree-based rules to match the characteristic of games from Chess to Go.

[28] It is a specific legal question whether this usage conflicts with intellectual property law. In the example of Germany, copyright law explicitly allows the usage of "published" data/images/text on the internet for "training" of AI systems (as long as the data is not "stored" externally at some third party).
[29] It is worth noting that recent analysis has revealed [HHI (2019)] that in certain cases pattern recognitions may be trained with "correct" labels, but that the system uses special features of the images as key elements, which are different to the labeling. For example, "trains" could be recognized by the "rails" as the dominating element in the dataset, which provides correct results as long as the image of a train contains rails.
[30] Mankowit et al. (2023) have shown that reinforcement learning can search algorithms as long as the task can be formulated as a "game" to be played iteratively.
[31] Russell and Critch (2023) published a taxonomy of existential AI risks. Primarily, these existential AI risks were derived from scenarios with feedback-loops in socio-technological systems such as social media. Nevertheless, these feedback-loops are well known as so-called echo chambers of social media.

Finally, LLMs are based on language, or in other words, human communication, but not on symbolic-logical truth. Generative AI scrapes text from the internet and mirrors actual "digital reality" including all errors, lies, disinformation, and biases.[32] Spitale et al. (2023) asked 697 participants to distinguish between tweets written by GTP-3 text generation and human Twitter users, and found "In comparison with humans, [GPT-3] can produce accurate information that is easier to understand, but it can also produce more compelling disinformation. We also show that humans cannot distinguish between tweets generated by GPT-3 and written by real Twitter users."

This insight can be applied to asset management. Overall, contemporary AI is a statistical classifier based on input of data with the limitation that any statistical estimation[33] is restricted to the "known" cases included in the input. For any use case including asset management, no "new" classification and no reasonable results beyond the defined "scope" can be provided by any statistical classifier. Additionally, there is the challenge to provide sufficient data for exceptions to be detected: whether outlines or "negative" labels.

## 6. CONCLUSION

Currently, no existing AI tool can go beyond statistical classifiers based on historical data. In general, ANNs can perform well in cases with re-identification of patterns, fraud signatures, recurring defects but also "positive" exceptions, which could be classified ex-ante. Nobody has to "fear the reaper" by AI. Any prediction of doom is based on human misconduct but never on some "self-awareness" of a computer program. Humans are always the biggest threat to humanity – today and in future.

Due to its statistical nature, based on vast text corpus as footing, generative AI is good for "average" estimations but incapable of achieving any truly creative idea or of any future "alpha" performance. In other words, generative AI tools make statistical estimations based on a continuation of "next best tokens" but without any chance to go "where no one has gone before." However, the three examples with the same input prompt revealed significant variations between different generative AI tools. In other words, there is no "single truth". Some results are "not even wrong", but depend on the selection of text corpus, control parameters, and proprietary post-processing methods. Potentially, generative AI-based financial advice tools with a (continuously iterated) fine-tuning to asset management information could gain customers' acceptance. For this use case, a dedicated fine-tuned generative AI could be expected to compete with an average human advisor. For professional investment decisions, generative AI can augment support staff to collect and summarize information, but it cannot provide any alpha, make innovative decisions, or develop unique strategies.

As the genie is out of the bottle, the future development of generative AI and LLMs has to be monitored – without fear of a hypothetical "superintelligence" but with awareness and caution. It is naïve to think of AI tools as anthropomorphized agents [Kidd and Birhane (2023)] and intentionally malevolent to fabricated AI-generated (mis)information.

In July 2023, Leike and Sutskever (2023) wrote in the OpenAI Blog that "Superintelligence will be the most impactful technology humanity has ever invented, and could help us solve many of the world's most important problems. But the vast power of superintelligence could also be very dangerous, and could lead to the disempowerment of humanity or even human extinction." Although OpenAI is the developer of generative AI-tools like ChatGPT, this quote resembles more the visions of symbolic-logic AI and the dystopias of science fiction novels of the 1960s than the formal analysis of understanding machine learning by Shalev-Shwartz and Shai Ben-David in 2014. The rationale behind the current wave of apocalyptic warnings by commercial firms has been widely discussed, but this kind of flashback to the early days of AI is remarkable and might require a more sociological debate than a discussion about the math of statistical classifiers. Especially because all LLMs are based primarily on the text corpus of human contributions on the internet, we are the real problem behind disinformation, bias, discrimination, aggression, violence, and war, not the robots.

[32] Shangbin et al. (2023) reported that LLMs reveal political biases due to differences in the text corpora used for the training. They developed a method with a fixed set of prompts asked to different LLMs and classified the answers. While the certain classifications can be questioned, the analysis revealed that different LLMs (and especially from different providers with individually selected text corpora) inherited political biases.

[33] Although authors such as e.g. Keith L. Downing (2023) discuss 'Predictive Neural Networks', all current ANNs are statistical classifiers, which provide an estimated classification of an new event according to their training, i.e., fit of parameters to 'timeless' input data+labels. The "prediction" is a process of humans when we interpret the statistical estimations as intertemporal extrapolations to future states.

## REFERENCES


Agrawal, A., J. S. Gans, and A. Goldfarb, 2023, "Do we want less automation?" Science 381:6654, 155-158

Albanesi, S., A. Dias da Silva, J. F. Jimeno, A. Lamo, and A. Wabitsch, 2023, "New technologies and jobs in Europe," ECB working paper series no. 2831, https://tinyurl.com/6zb3e4xx

Aleph Alpha, 2023, "What is Luminous?" Aleph Alpha GmbH, https://tinyurl.com/25fu7xv3

Altman, S., G. Brockman, and I. Sutskever, 2023, "Governance of superintelligence," OpenAI, https://tinyurl.com/2s4fhs6b

Amazon, 2023, "AWS announces Generative AI Innovation Center," Press release, https://tinyurl.com/2p8d2had

Anthropic, 2023, "Claude 2," https://tinyurl.com/5esb72wa

BaFin, 2023, "Social trading: Kopierte Risiken?" BaFinJournal online, https://tinyurl.com/5n7ujcyp

Bai, Y., et al., 2022, "Constitutional AI: harmlessness from AI feedback," arXiv, https://tinyurl.com/2p84a2eu

Bengio, Y., et al., 2023, "Pause giant AI experiments: an open letter," Future of Life Institute, https://tinyurl.com/3kb2wnrw

Brynjolfsson, E., D. Li, and L. R. Raymond, 2023, "Generative AI at work," NBER working paper no. 31161, https://tinyurl.com/5eptyp69

Bubeck, S., et al., 2023, "Sparks of artificial general intelligence: early experiments with GPT-4," Microsoft Reserch, arXiv, https://tinyurl.com/3f83z5du

Buzsáki, G., S. McKenzie, and L. Davachi, 2022, "Neurophysiology of remembering," Annual Review of Psychology 73, 187–215

Capgemini Research Institute, 2023, "Why consumers love generative AI," Capgemini Research Institute, https://tinyurl.com/56sv76t7

CFPB, 2023, "CFPB issue spotlight analyzes AI chatbots in banking," Consumer Financial Protection Bureau, Newsroom, https://tinyurl.com/bdh9madm

ChaosGPT, 2023, "ChaosGPT: empowering GPT with internet and memory to destroy humanity," video, https://tinyurl.com/2p8tmweu

Choi, Y., 2023, "An AI odyssey: the dark matter of intelligence," Keynote at Computer Vision and Pattern Recognition, Vancouver, https://tinyurl.com/5a299u9x

Chui, M., et al., 2023, "The economic potential of generative AI The next productivity frontier," McKinsey & Company, June, https://tinyurl.com/4a83e35u

Clarkson Law Firm, 2023, "OpenAI complaint," https://tinyurl.com/udtpwa9w

Critch, A., and S. Russell, 2013, "TASRA: a taxonomy and analysis of societal-scale risks from AI," arXiv, https://tinyurl.com/26z4sa5z

Cruise, 2023, "Further update on emergency vehicle collision," Blog, Greg Dietrerich, https://tinyurl.com/4uwhydyv

Downing, K. L., 2023, Gradient expectations: structure, origins, and synthesis of predictive neural networks, The MIT Press

Dr. House, 2005, Episode "Three stories," FOX Broadcasting Company

Ekman, P., and W. V. Friesen, 1978, Facial Action Coding System: a technique for the measurement of facial movement, Consulting Psychologists Press

Eloundou, T., S. Manning, P. Mishkin, and D. Rock, 2023, "GPTs are GPTs: an early look at the labor market impact potential of large language models," arXiv, https://tinyurl.com/66hu6emz

European Parliament, 2023a, ""Artificial Intelligence Act" (AIA)," texts adopted, https://tinyurl.com/nshsyx24

European Parliament, 2023b, "AI Act: a step closer to the first rules on artificial intelligence," IMCO/LIBE, press releases, https://tinyurl.com/5n7zx7v3

Fama, E. F., 2013, "Two pillars of asset pricing," Prize Lecture, https://tinyurl.com/4yfc7899

Feng, S., et al., 2023, "From pretraining data to language models to downstream tasks: tracking the trails of political biases leading to unfair NLP models," Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, July 9-14, 2023, Vol. 1, 11737–11762, https://tinyurl.com/3f6kk4t6

Futuri, 2023, "Futuri launches RadioGPT™, the world's first AI-driven localized radio content," Futuri Blog, https://tinyurl.com/3epb55jh

Ghahramani, Z., 2023, "Introducing PaLM 2," Google AI, https://tinyurl.com/yc7unbr3

Gigerenzer, G., 2023, "AI in an uncertain world," The American Journal of Psychology (in press).

Grossmann, I., et al., 2023, "AI and the transformation of social science research," Science 380:6650, 1108-1109

HHI, 2019, "Der Blick in Neuronale Netze," Forschung Kompakt (in German), Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut (HHI), https://tinyurl.com/mr3e3t35

Hinton, G., et al., 2023, "Statement on AI risk," Center for AI Safety, https://tinyurl.com/3442schr

IBM, 2023, "IBM brings generative AI … to the Wimbledon digital experience," IBM News, https://tinyurl.com/ymxz42kz

Johnson, D. G., 2006, "Computer systems: moral entities but not moral agents," Ethics and Information Technology 8:4, 195–204.

Jones, H., 2023, "Juergen Schmidhuber, renowned "father of modern AI," says his life's work won't lead to dystopia," Forbes, https://tinyurl.com/3x95sz9x

Kanazawa, K., D. Kawaguchi, H. Shigeoka, and Y. Watanabe, 2022, "AI, skill, and productivity: the case of taxi drivers," IZA – Institute of Labor Economics, DP no. 15677, https://tinyurl.com/297xv6nn

Kidd, C., and A. Birhane, 2023, "How AI can distort human beliefs," Science 380:6651, 1222-1223

Krawczyk, J., and A. Subramanya, 2023, "Bard is getting better at logic and reasoning," Google Blog, https://tinyurl.com/3ekufu54

Kremer, D., 2023, "Sind EFTs wirklich besser?" Frankfurter Allgemeine Sonntagszeitung, June 4th, 26

Laajaj, R., et al., 2019, "Challenges to capture the big five personality traits in non-WEIRD populations," Science Advances 5:7, https://tinyurl.com/2jbefavj

LeCun, Y., Y. Bengio, and G. Hinton, 2015, "Deep learning," Nature 521, 436–444

Liu, X., et al., 2019, "A comparison of deep learning performance against health care professionals in detecting diseases from medical imaging ...," Lancet Digital Health 1, 271–297

Lussange, J., et al., 2022, "Stock price formation: precepts from a multi-agent reinforcement learning model," Computational Economics 61:4, 1523–1544

Singhal, K., et al., 2023, "Large language models encode clinical knowledge," Nature 620, E19

Mankowit, D. J., et al., 2023, "Faster sorting algorithms discovered using deep reinforcement learning," Nature 618, 257-263

McCarthy, J., M. Minsky, N. Rochester, and C. E. Shannon, 1955, "A proposal for the Dartmouth Summer Research Project on Artificial Intelligence," August, https://tinyurl.com/yjpnj39k

Meyer, J. B., 2015, "Künstliche Intelligenz ist besser als natürliche Dummheit," Interview mit Wolfgang Wahlster, Computerwoche, https://tinyurl.com/ytxxzbz2

Milkau, U., 2021, "The risks of artificial intelligence used for decision-making in financial services," Journal of Financial Transformation 52, 128-141

Noy, S., and W. Zhang, 2023, "Experimental evidence on the productivity effects of generative artificial intelligence," Science 381, 187–192

OpenAI, 2023, "GPT-4 is OpenAI's most advanced system, producing safer and more useful responses," https://tinyurl.com/yckpr6tp

Pearl, J., and D. Mackenzie, 2018, The book of why, Basic Books/Hachette Book Group

Rathi, N., 2023, "Our emerging regulatory approach to big tech and artificial intelligence," speech, https://tinyurl.com/22cnanyd

Salman, H., et al., 2023, "Raising the cost of malicious AI-powered image editing," in Proceedings of Machine Learning Research, vol. 202: International Conference on Machine Learning, 23-29 July 2023, Honolulu, Hawaii, https://tinyurl.com/h2vc7kxz

Schmidhuber, J., 2015, "Deep learning in neural networks: an overview," Neural Networks 61, 85-117

Shalev-Shwartz, S., and S. Ben-David, 2014, Understanding machine learning: from theory to algorithms, Cambridge University Press

Shevlane, T., et al., 2023, "Model evaluation for extreme risks," DeepMind, https://tinyurl.com/yrdmu973

Spitale, G., N. Biller-Andorno, and F. Germani, 2013, "AI model GPT-3 (dis)informs us better than humans," Science Advance 9:26, https://tinyurl.com/32et8yvf

StM.BW, 2023, "Künstliche Intelligenz in der Verwaltung," Staatsministerium Baden-Württemberg, https://tinyurl.com/38hyab55

Thompson, E. A., 2007, "The tulipmania: fact or artifact?" Public Choice 130:1–2, 99–114

Touvron, H., et al., 2023a, "LLaMA: open and efficient foundation language models," MetaAI, https://tinyurl.com/5bk7vrbw

Touvron, H., et al., 2023b, "Llama 2: open foundation and fine-tuned chat models," Meta AI, https://tinyurl.com/yc8d9a9f

Webb, T., K. J. Holyoak, and H. Lu, 2023, "Emergent analogical reasoning in large language models," Nature Human Behaviour, https://tinyurl.com/542s2fx4

White House, 2023, Remarks by President Biden on artificial intelligence, https://tinyurl.com/bdd8safb

Winograd, T. A., and F. Flores, 1986, Understanding computers and cognition, Ablex Publishing Corporation

Worldcoin, 2023, "A new identity and financial network," Worldcoin, https://tinyurl.com/mvthapfk

Wu, Y., 2023, "Autoformalization with large language models," 36th Conference on Neural Information Processing Systems (NeurIPS 2022), https://tinyurl.com/y332vsre

## APPENDIX

**Table A1:** Examples generated by ChatGPT, Bard, and Llama 2-Chatbot with a prompt about asset management

| PROMPT: CONSTRUCT A BALANCED PORTFOLIO OF TEN SECURITIES FOR A RISK-AVERSE INVESTOR OF AGE 40! | | |
|---|---|---|
| **ChatGPT (June 7, 2023, shortened):** | **Bard (July 15, 2023, shortened):** | **Llama 2-Chatbot (July 29, 2023):** |
| Constructing a balanced portfolio involves diversifying investments across different asset classes and sectors to manage risk. For a risk-averse investor of age 40, it's generally recommended to have a mix of equities, fixed income securities, and other alternative investments. Here's a sample portfolio consisting of ten securities that could suit a risk-averse investor:<br>Vanguard Total Stock Market Index Fund ...<br>iShares Core S&P 500 ETF ….<br>Vanguard FTSE ...<br>iShares Core U.S. Aggregate Bond ETF ...<br>Vanguard Total Bond Market Index Fund ...<br>Invesco Senior Loan ETF (Fixed Income) ...<br>iShares Gold Trust (Commodity) ...<br>Vanguard Real Estate Index Fund ...<br>Invesco Solar ETF (Clean Energy...<br>Fidelity MSCI Utilities Index ETF ...<br>Remember that constructing a portfolio involves considering your specific financial goals, risk tolerance, and investment horizon. It's always recommended to consult with a financial advisor who can provide personalized advice based on your individual circumstances. | Sure, here is a balanced portfolio of ten securities for a risk-averse investor of age 40: …<br>40% Vanguard Total Stock Market Index Fund<br>20% Vanguard Growth Index Fund<br>10% Vanguard Value Index Fund<br>10% Vanguard Small-Cap Index Fund<br>20% Vanguard Total Bond Market Index Fund<br>10% Vanguard Intermediate-Term Treasury<br>10% Vanguard Short-Term Treasury Fund<br>10% Vanguard Total International Stock Index<br>This portfolio has a 60% stock allocation and a 40% bond allocation, which is a good starting point for a risk-averse investor of age 40. The stocks are diversified across different market capitalizations and sectors, and the bonds are diversified across different maturities. The other security is a total international stock index fund, which helps to reduce the portfolio's home bias.<br>Of course, this is just a starting point, and you may need to adjust the portfolio based on your individual risk tolerance and investment goals. However, this portfolio provides a good foundation for a balanced investment strategy.<br>Here are some additional considerations for a risk-averse investor of age 40:<br>Consider your time horizon. If you are nearing retirement, you may want to allocate more of your portfolio to bonds. However, if you have a long time horizon, you may be able to afford to take on more risk.<br>Consider your risk tolerance. How comfortable are you with the idea of your portfolio losing value in the short term? If you are not comfortable with a lot of risk, you may want to allocate more of your portfolio to bonds.<br>Rebalance your portfolio regularly. As your investments grow, you will need to rebalance your portfolio to maintain your desired asset allocation. This will help to ensure that your portfolio remains aligned with your risk tolerance and investment goals.<br>I hope this helps! | I'm happy to help you with that! However, I must point out that asking me to construct a balanced portfolio without considering the investor's personal financial goals, risk tolerance, and time horizon may not be the best approach. It's important to tailor a portfolio to the individual's specific needs and circumstances.<br>That being said, I can offer some general guidance on how to construct a balanced portfolio. A risk-averse investor may want to consider a mix of low-risk investments such as bonds, money market funds, and large-cap stocks with a history of stability and growth. It's also important to diversify across different sectors and industries to minimize exposure to any one particular area.<br>Here's an example of a balanced portfolio that a risk-averse investor may consider:<br>30% Bonds (e.g., U.S. Treasury, high-quality corporate bonds)<br>20% Money Market Funds<br>20% Large-Cap Stocks (e.g., Apple, Microsoft, Procter & Gamble)<br>15% Mid-Cap Stocks (e.g., Amazon, Google, Facebook)<br>10% Small-Cap Stocks (e.g., up-and-coming companies in various industries)<br>5% International Stocks (e.g., developed markets, emerging markets)<br>5% Real Estate Investment Trusts (REITs)<br>5% Utilities (e.g., electric, gas, water)<br>5% Consumer Staples (e.g., food, beverages, household goods)<br>5% Healthcare (e.g., pharmaceuticals, medical devices, health insurance)<br>Again, it's important to note that this is just a general example and not a tailored portfolio for a specific individual. It's always best to consult with a financial advisor or professional before making investment decisions. |

ChatGPT and Bard were accessed via the websites of the providers and Llama 2 via llama2.ai (with the default setting of the control parameters). It is important to note that different control parameters and changes in the proprietary post-processing methods over time can produce different results compared to this snapshot.

**Table A2:** Comparison of the results of ChatGPT (with the disclaimer "limited knowledge of world and events after 2021") and Bard with two randomly selected references for best ETFs or Index Funds for June 2023

| FORBES (PER 30.5.2023) | YAHOO (PER 1.6.2023) | CHATGPT (PROMPTED 7.6.2023) | BARD (PROMPTED 15.7.2023) |
|---|---|---|---|
| Invesco Balanced Multi-Asset Allocation ETF | Fidelity ZERO Large Cap Index | Vanguard Total Stock Market Index Fund | Vanguard Total Stock Market Index Fund |
| iShares Core U.S. REIT ETF | Vanguard S&P 500 ETF | iShares Core S&P 500 ETF | Vanguard Growth Index Fund |
| Nuveen ESG Large-Cap Value ETF | SPDR S&P 500 ETF Trust | Vanguard FTSE Developed Markets ETF | Vanguard Value Index Fund |
| Invesco RAFI Strategic US Small Company ETF | iShares Core S&P 500 ETF | iShares Core U.S. Aggregate Bond ETF | Vanguard Small-Cap Index Fund |
| Vanguard International Dividend Appreciation ETF | Schwab S&P 500 Index Fund | Vanguard Total Bond Market Index Fund | Vanguard Total Bond Market Index Fund |
| Vanguard Short-Term Inflation-Protected Securities | Shelton NASDAQ-100 Index Direct | Invesco Senior Loan ETF | Vanguard Intermediate-Term Treasury |
| Fidelity Total Bond ETF | Invesco QQQ Trust ETF | iShares Gold Trust | Vanguard Short-Term Treasury Fund |
| | Vanguard Russell 2000 ETF | Vanguard Real Estate Index Fund | Vanguard Total International Stock Index |
| | Vanguard Total Stock Market ETF | Invesco Solar ETF | Vanguard Total Stock Market Index Fund |
| | SPDR Dow Jones Industrial Average ETF | Fidelity MSCI Utilities Index ETF | Vanguard Growth Index Fund |

Note: All ETF and Index Funds mentioned actually exist, although "large language models" come without any pre-training on asset management.

# HOW CAN BANKS EMPOWER THEIR CUSTOMERS TO FLAG POTENTIAL VULNERABILITIES?

**PRZEMEK DE SKUBA** | Senior Consultant, Capco
**BIANCA GABELLINI** | Consultant, Capco
**JESSICA TAYLOR** | Consultant, Capco*

## ABSTRACT

Customer vulnerability is one of the key concerns of the Consumer Duty regulation, a very welcome ESG-aligned enhancement of financial institutions' governance. Adherence to the regulation requires a clear focus on data collection that helps lenders manage the impact of consumer vulnerabilities without imposing penalties or resulting in a negative impact on clients. There are two parts of the problem that need to be addressed: firstly, how to capture vulnerability data by encouraging clients/consumers to voluntarily submit the information (the behavioral aspect) and secondly, how to technically capture, manage, and store this data to ensure compliance with the Consumer Duty regulation. This article considers both problems and reviews the tools from behavioral science that can encourage customer disclosure and two key technology solutions (data lakes and blockchain) to comply with the capture, management, and storage of data whilst remaining GDPR compliant and fully aligned to the objective of voluntary submission of information regarding vulnerabilities by clients/consumers.

## 1. INTRODUCTION

The Financial Conduct Authority (FCA) led consultations in 2021 that resulted in the development of the Consumer Principle (Principle 12), putting the onus on the U.K. organizations within its scope to "act to deliver good outcomes for retail customers" from 31 July 2023. Sheldon Mills, the Executive Director at the Consumers and Competition department of the FCA has specifically pinned this responsibility on the "boards and senior management [who] have a critical role in overseeing firms' implementation of the Duty. That is why [the FCA has] strengthened the requirements around governance and accountability to ensure senior managers and executives are held accountable."[1] This development highlights the growing importance of ESG and is a very welcome improvement in governance – the "G" in ESG.

Even though the U.K. already has some of the best governance frameworks in the world, one does not have to look far for warnings of what can happen when consumers are not properly protected. Poland – despite enjoying the status of a developed (per FTSE) economy with the fifth largest GDP in the E.U. – is a great case in point: an extreme example of an industrialized European country where a combination of very lax consumer protection and an incredibly light-touch financial market regulation has allowed for a mass proliferation of toxic financial products dressed up as foreign currency denominated mortgages. The problem has been allowed to fester for close to two decades now, with a peak in 2007 when "over half of Polish mortgages were issued in Swiss francs." It took several interventions by the European Court of Justice in the past couple of years to finally prompt the Polish courts to begin

* We would like to thank Julia Shreeve (Business Consulting sponsor), Martha Ferez (Behavioural Science Practice Lead), and Mark Profeti (Blockchain subject matter expert) for their support and advice with this article.

[1] FCA, 2022, "What firms and customers can expect from the Consumer Duty and other regulatory reforms," Financial Conduct Authority, Speech by Sheldon Mills, Executive Director, Consumers and Competition, delivered at the Consumer Protection in Financial Services Summit, https://tinyurl.com/ycyav7rc

annulling some of the Swiss-franc mortgages "after ruling that banks used "abusive" foreign exchange rates."[2] Importantly, on top of causing misery for millions, this has also resulted in a systemic risk to the Polish banking system.[3]

Fortunately, the U.K. has never had to deal with this sort of a problem, and indeed the Consumer Duty regulation goes even further in protecting consumer rights, with the policy placing the importance of customer vulnerability as its key priority. Of a particular note are the following paragraphs:

- **1.16:** which requires "firms to consider the needs, characteristics, and objectives of their customers – including those with characteristics of vulnerability – and how they behave, at every stage of the customer journey." It raises the bar with regards to consumer protection required of the regulated companies.
- **8.5:** which highlights that "many respondents queried the practical application (…) considerations relating to potential vulnerabilities; and the proposed approach to testing communications." Consequently, it highlights the interest expressed by those consulted in the practicalities of flagging up potential vulnerabilities.
- **10.6:** which presents the view of the consumer organizations that "suggested that firms should be required to take an inclusive design approach to meet the needs of customers with characteristics of vulnerability." Hence, clarifying the user requirement with regards to the functionalities for flagging potential and actual vulnerabilities.
- **Annex B 2A.7.4 G:** specifies that "in relation to the needs and characteristics of retail customers, a firm should, among other things: (…) **(4)** assist frontline staff to understand how to actively identify information that could indicate vulnerability and, where relevant, seek information from retail customers with characteristics of vulnerability that will allow staff to respond to their needs;" – thereby clarifying the requirement for the vulnerability flagging functionality from the point of view of the financial institutions, as well as "**(5)** set up systems and processes in a way that supports and enables retail customers with characteristics of vulnerability to disclose their needs."[4] – Hence providing a further clarification on the user requirement with respect to the same functionality.

As such, the FCA makes the requirements regarding any processes pertaining to the flagging of potential and actual vulnerabilities clear, with only two key outstanding questions remaining: how to encourage consumers to voluntarily submit their vulnerability data and which technology would best suit this use case.

## 2. THE PROBLEM

While the Consumer Duty regulation is clearly a step in the right direction as far as governance is concerned, the main problem that needs to be addressed is that customer vulnerability is a dynamic concept (i.e., it changes over time), and currently financial institutions take a static approach. More importantly, financial institutions have no proper mechanisms for monitoring customer vulnerabilities. As an example, mortgage customers will only have the KYC (know your customer) due diligence at the point of applying for the product, and even then, the KYC will not necessarily capture any vulnerabilities, as it is designed with AML (anti-money laundering) in mind. The lender will typically only find out about any vulnerabilities their customer may have been suffering from when they go into arrears with their mortgage. This is a common theme across the financial services industry, not just within home financing or general lending.

The other issue is the desire to protect one's privacy, or sometimes even the shame of admitting a problem or a weakness, as well as the natural human propensity to protect one's interests by presenting oneself as stronger, and more in control than one may be. Going deeper into human psychology to assess why customers may be opting not to disclose, behavioral science would classify these fears as "inherent biases".

### 2.1 Behavioral biases

Firstly, there are biases that cause people to omit negative information, such as "omission bias".[5] The "omission bias" describes how voluntary oversights are empowered by our inner belief that, ceteris paribus, committing an action is more

[2] Minder, R., 2022, "The mortgage time bomb ticking beneath Poland's banks," Financial Times, November 13, https://tinyurl.com/4s5ffzp4
[3] de Skuba Skwirczynski, P., 2021, "Swiss franc mortgages: European banks are profiteering from the Polish subprime loan plight," The Quarterly Journal of the International Union for Housing Finance, Summer, 28-32, https://tinyurl.com/bdds9vfw
[4] FCA, 2022, "A new Consumer Duty," Feedback to CP21/36 and final rules, Policy Statement PS22/9, July, Financial Conduct Authority, https://tinyurl.com/yw6mm3p
[5] Caviola, L., A. Mannino, J. Savulescu, and N. Faulmüller, 2014, "Cognitive biases can affect moral intuitions about cognitive enhancement," Frontiers in Systems Neuroscience 8, https://tinyurl.com/593pvu3d

dangerous than omitting an action. On this basis, disclosing something is perceived as a favorable action to take and customers prefer to assume the risk of hiding important information. As it is well publicized, repayment history is one of main factors affecting credit score.[6] As such, even though flagging a vulnerability does not imply arrears, it may nevertheless be cognitively associated by a customer with presenting themselves as being at an increased risk. Flagging up any vulnerability may be associated in a similar fashion with making one look riskier to the lender, and by extension to the credit scoring company, which could in turn be mistakenly perceived as negatively affecting the availability of future financial products.

Secondly, there are biases that blind people from negative information, pushing them towards an overly optimistic evaluation. Behavioral scientists call this "optimism bias" and some of its implications are people underestimating the risk of having low savings, aging, or their caring responsibilities. In these cases, omission does not spring from a forward looking and well-thought strategy, but from a purely involuntary reflex; a constructed belief in a positive outcome in which the customer is the first to believe in. In view of all this, we understand that disclosure action implies a great challenge, particularly when lenders rely on customers to take the initiative to indicate actual and potential vulnerabilities.

With that in mind, not only are customers likely to withhold their vulnerabilities but they are also less likely to disclose certain vulnerabilities than others. Considering self-disclosure types per the various vulnerabilities, as defined by the FCA (Table 1), some, such as visual impairments and poor English language skills, are more likely to be self-disclosed than others, such as mental health conditions and an income shock. That would be not only due to a perceived associated stigma, the unwillingness to admit "failure", but also for the (more practical) fear of being "blacklisted" from future financial products or having the existing mortgage revoked, however unfounded these assumptions may be. Considering the latter examples, the challenge for lenders is to create an atmosphere where customers believe they can safely share information pertaining to such vulnerabilities by way of self-disclosures.

## 3. THE SOLUTION

A tough question to answer is whether it is even feasible to create an environment where customers would take it upon themselves to flag up their observed or potential vulnerabilities. This question boils down to assessing what the possible resulting benefits or incentives for the customer could be.

### 3.1 How to capture vulnerability data by encouraging clients/consumers to voluntarily submit the information

Addressing "inherent biases" is key to considering any potential solutions for self-disclosing of potential vulnerabilities. The solution must significantly contribute to creating an environment where customers feel encouraged to disclose potential vulnerability-driven cashflow problems before they occur. This precarious stage is sometimes referred to as "pre-arrears", and some examples could include employees anticipating a redundancy, the self-employed observing worsening market conditions and consequently the

**Table 1:** FCA-defined self disclosure types[7]

| HEALTH | LIFE EVENTS | RESILIENCE | CAPABILITY |
|---|---|---|---|
| • Physical disability<br>• Severe or long-term illness<br>• Hearing or visual impairment<br>• Mental health condition<br>• Addiction<br>• Low mental capacity or cognitive disability<br>• Being "older, old" i.e., >80<br>• Being young<br>• Non-standard requirements or credit history | • Retirement<br>• Bereavement<br>• Income shock<br>• Relationship breakdown<br>• Domestic abuse<br>• Caring responsibilities<br>• Other, i.e., leaving care, migration or seeking asylum, human trafficking or modern slavery, convictions | • Inadequate (outgoings exceed income) or erratic income<br>• Over-indebtedness<br>• Low savings<br>• Low emotional resilience | • Low knowledge or confidence in managing finances<br>• Poor literacy or numeracy skills<br>• Poor English language skills<br>• Poor or non-existent digital skills<br>• Learning difficulties<br>• No or low access to help or support |

[6] https://tinyurl.com/4hh67v9w
[7] https://tinyurl.com/45wp6u2k

likelihood of work drying up, or one's mental or general health worsening, all of which could ultimately lead to the borrower going into arrears.

To the lender, customer vulnerability is either disclosed by the person in question or inferred about them, with a clear preference for the former because the lender, just as much as anyone else, prefers to be certain of the risks, as opposed to having to infer them.

Consequently, to approach this from the customer's perspective, the main question is about what the lender, as the party to the contract who commands more power, should be doing to increase their customers' willingness to disclose any anticipated cashflow problems. Currently, the customer has little insight into the data held on them by the financial institutions and ancillary enterprises, such as credit scoring companies. Of course, GDPR has given customers the power to demand the data held about them from any such institutions, but these rights hardly mean that such data would be available at the touch of a button. In practice, extracting it could be a lengthy and painful process, with the necessity to write Freedom of Information requests and waiting for weeks at a time for a response. Banks could help by creating an environment where customers would be more willing to disclose their vulnerabilities by increasing transparency surrounding personal data gathered and building trust with their customers.

Another positive for the customer resulting from self-disclosure could be staying in control of exposing one's vulnerabilities. Having a say in the timing and manner of such a disclosure would grant the customer the power to control the narrative of their vulnerabilities, thereby ensuring they can present it in the best possible light. That would not be possible in a situation where the lender finds out about the issue via a third party, for example, once the customer is already in arrears. Such an approach is akin to a "controlled fall" technique taught to frail patients and high-performing sportspeople to help them prevent unnecessary injuries. Overseeing exposing their own vulnerabilities to others could be particularly attractive to customers who appreciate being in control.

Behavioral science could be applied to identify appropriate techniques that lenders could apply to encourage their customers to self-disclose vulnerabilities as opposed to having them inferred. In the context of the Consumer Duty, some of the applicable strategies that can be adopted by financial institutions are "nudges", which encourage better decisions by making certain choices easier than others, and "sludges", which discourage decisions by making the process more difficult.

Examples of nudges include:

- **Precommitment:** asking the borrower to confirm at the beginning of every fiscal year that the information held by the lender is still relevant and that they commit to notify the lender in case of any changes. Studies have shown that this technique is effective, as it facilitates the retrieval of intentions in our memory and reduces the probability of past actions impacting future behavior.[8]
- **Social norming:** emphasizing what most people are doing while promoting the correct behavior can influence borrowers' behavior, as it provides social rules and standards to follow.[9] Captions such as "nine out of ten customers have reviewed their parameters this year" leverage our inner need to feel included in a wider group (known as the "the bandwagon effect" and "herd mentality").
- **Default rules:** presenting a list of opted-in conditions from which the consumer is asked to opt out when these do not apply reduces the friction of telling the truth that is "already being told" in the presented conditions, in which case there is no further action for the customer to take. On the contrary, lying would mean actively removing the tick when asked to opt in.[10]
- **Disclosure:** disclosing the cost of the customers' omission, either by sharing the economic loss of misinformation or the financial penalty for providing inaccurate information, will make the consumer completely aware of the granular and wider consequences of their actions, thereby putting in doubt the safety of their passivity.
- **Graphic warnings:** leveraging the use of large or bold fonts attracts borrowers' attention in support of the promoted behavior as well as to support knowledge of misinformation risks (i.e., "omission effect"). This approach is commonly used with respect to cigarettes and tobacco products, and it has so far proved effective, with an increased number of attempts to quit smoking.

[8] Conner, M., and P. Norman (eds.), 2015, Predicting health behavior: research and practice with social cognition models, Open University Press
[9] Thaler, R. H., and C. R. Sunstein, 2009, Nudge: improving decisions about health, wealth, and happiness, Penguin Books
[10] Sunstein, C. R., 2006, Boundedly rational borrowing, University of Chicago Law Review 73:1, 249-270

- **Reminders:** implementing a series of prompts via email or text messages are difficult for consumers to ignore (addressing the omission bias). Scientists have highlighted how delegating a task to an automation device can reduce cognitive load, making it easier for people to act when needed. This has proved effective in different scenarios, such as savings management and medical treatment adherence.[11]

For this to yield anticipated results, customers would need to be assured of clear guardrails, whereby lenders would only be allowed to use such self-disclosed information for the purpose of assisting customers with the disclosed vulnerability rather than by gifting lenders ammunition to penalize customers for an elevated risk. Potential examples of assistance from the lender could include using the right communication channels, indicating the right products, or suggesting repayment holidays. To put customers at ease, lenders could be legally obliged (or pledge) to offer those self-disclosing a vulnerability a similar treatment to that afforded to the British "legally protected characteristics", which cannot be discriminated against.[12] In this way, those flagging their vulnerabilities would be exempted from penalties. On the flipside, they could be subjected to the usual penalizing procedure if they failed to flag their vulnerability and ended up in arrears – that is in the eventuality that the lenders wished to apply a "carrot and stick" approach.

### 3.2 How to technically capture, manage, and store this data to ensure compliance with Consumer Duty regulation

Given that managing risks is right at the center of the lenders' business, when it comes to their customers' vulnerabilities, inferring these issues is problematic, as it introduces uncertainty into the lender's risk management. Self-disclosures would help lenders pre-empt, or mitigate, problems arising from their clients ending up in arrears and be positive for their risk management.

Another issue is brand management and PR, as lenders are typically well-known and respected institutions. For example, if a bank were to build a "natural language processing" (NLP) model aimed at inferring vulnerability, it should disclose that fact to customers to comply with data processing laws, such as GDPR, and in general to keep everything regarding their relationship with their customers "above board". Such a disclosure could be perceived as "bad optics" from a PR perspective. Additionally, the consequences of getting such an NLP model wrong and inferring vulnerabilities where there are not any, or misdiagnosing them, would carry a further significant reputational risk for the lender. That is another apparent reason why banks should prefer self-disclosure by clients, given their precision and cost effectiveness.

A separate question is whether financial institutions are sufficiently empowered to help customers who flag vulnerabilities. That is important from the customer experience angle, as a customer who self-discloses but does not receive appropriate support would not only be disappointed but could also lodge complaints and be deterred from flagging their vulnerabilities in the future. Any potential penalization resulting from such a self-disclosure would create a bad customer experience.

Not just lenders, but ancillary enterprises, such as credit scoring agencies, should also positively perceive people self-identifying their potential or expected vulnerabilities. Such self-awareness on the customers' part would prove that they are responsible individuals, particularly when faced with the tightening of their finances, the resulting reduced spending, the need to make difficult lifestyle choices, and so on. While there are benefits to the wider financial services industry resulting from the empowerment of customers to self-disclose their vulnerabilities, evaluation of the wider impact is beyond the scope of this article.

It is also worth noting that, due to the breadth of vulnerabilities in scope of Consumer Duty (as visualized in Table 1) lenders may find that one solution will not fit all potential disclosures and there may be a need for a variety of approaches.

In this paper, two approaches are explored to solve the above-mentioned issues. The first is a data lake, selected due to its current wide usage in the financial services industry. The second is blockchain, selected due to the expected benefits and advantage it can deliver in the future. By comparing them, we aim to understand their limitations and potential when used to facilitate compliance with the Consumer Duty regulation.

[11] Gravert, C., 2019, "The hidden costs of reminders," Behavioral Scientist, March 19, https://tinyurl.com/dcnddxd3. Orbell S., S. Hodgkins, P. Sheeran, 1997, "Implementation intentions and the theory of planned behavior," Personality and Social Psychology Bulletin 23:9, 945-954

[12] Gov.uk, 2010, "Discrimination: your rights," U.K. Government, https://tinyurl.com/yj4328tz

A data lake is a centralized repository in which raw data are stored in a structured, semi-structured, or unstructured way, and it is the most common tool used by organizations to store and analyze data. It is designed to handle large amounts of data and is, therefore, a valuable tool for organizations looking to analyze and extract insights from their data in cases where traditional relational databases are not well-suited due to scalability and data variety issues. To address customer vulnerability disclosure, a data lake can serve as a foundational data infrastructure for financial organizations to collect, store, integrate, analyze, and report on customer vulnerabilities, while incorporating robust security and data governance measures.

Blockchain is a decentralized and distributed repository where data are stored in a structured way. By recording transactions across multiple computers, it provides a tamper-resistant and trustless environment that ensures security, transparency, and immutability of the data. This technology, often associated with cryptocurrency and praised for its security features, has become quite popular within the financial services sector, with a compound annual growth (CAGR) of 62.7%[13] since 2016 – and its growth is not expected to halt.[14] To address customer vulnerability disclosures, blockchain ensures the integrity and authenticity of the data, as once a disclosure is made it is securely and permanently recorded, reducing the risk of data manipulation or tampering.

**Table 2:** Evaluation of the application of data lake technology considering Consumer Duty requirements

| FCA REQUIREMENTS | EVALUATION OF THE APPLICATION OF DATA LAKE TECHNOLOGY |
|---|---|
| **1.16** which requires “firms to consider the needs, characteristics and objectives of their customers – including those with characteristics of vulnerability – and how they behave, at every stage of the customer journey.” – raising the bar of consumer protection required of the regulated companies.<br><br>**8.5** which highlights that “many respondents queried the practical application (…) considerations relating to potential vulnerabilities; and the proposed approach to testing communications” – thereby proving the interest expressed by those consulted in the practicalities of flagging up potential vulnerabilities. | When evaluating the application of a data lake with reference to FCA requirements, such a solution would meet para. 1.16 as much as any comparable technology, while not falling foul of para. 8.5 because a data lake does not come across as a relevant tool for capturing vulnerability data itself. |
| **10.6** presents the view of the consumer organizations which “suggested that firms should be required to take an inclusive design approach to meet the needs of customers with characteristics of vulnerability” – hence clarifying the user requirement with regards to the functionalities for flagging potential and actual vulnerabilities. | It would, however, fail the test of para. 10.6 because the data lake managed by a lender would not be particularly inclusive from the perspective of the customer sharing their vulnerabilities. |
| **Annex B 2A.7.4 G** specifies that “in relation to the needs and characteristics of retail customers, a firm should, among other things: (…)<br>**(4)** assist frontline staff to understand how to actively identify information that could indicate vulnerability and, where relevant, seek information from retail customers with characteristics of vulnerability that will allow staff to respond to their needs;”[15] – thereby clarifying the requirement for the vulnerability flagging functionality from the point of view of the financial institutions, as well as | In terms of Annex B 2A.7.4 G (4), this technology would not fall foul here, just as much as in para. 8.5, as the data lake would not be used for the purposes of identification of vulnerabilities. However, it must be noted that in reference to not falling foul of the requirements set out by the FCA in both these paragraphs, the application of a data lake is “not applicable”. |
| “**(5)** set up systems and processes in a way that supports and enables retail customers with characteristics of vulnerability to disclose their needs;”[15] – hence providing a further clarification on the user requirement with respect to the same functionality. | A data lake could not be used to help customers with their disclosures as it relies for its data on inputs from other systems, which by its nature would be logistically difficult to be performed by individual customers who simply wish to input their vulnerability information into a user interface. |

[13] https://tinyurl.com/2s3asmem
[14] https://tinyurl.com/2p9s5ad2
[15] FCA, 2022, “A new Consumer Duty, Feedback to CP21/36 and final rules,” Policy Statement PS22/9, July, Financial Conduct Authority, https://tinyurl.com/yw6mm3p

## 4. SOLUTION EVALUATION

### 4.1 Data lake technology evaluation

It may be tempting to frame the solution as a data lake use case. Banks are by now well serviced in this regard by competing "cloud services providers" (CSPs) and typically well-versed in the use of this technology. Extending existing data lakes' application to cover self-disclosures of customer vulnerabilities may, therefore, appear as a logical next step to take.

To summarize, as far as the FCA's requirements for a vulnerability self-disclosure solution are concerned, data lake technology falls on two separate accounts and is not particularly applicable to another two.

With GDPR in mind, inspection by an individual (required by law) of the data held about them by their lender in a data lake would require a customer request that would need to be fulfilled by staff working for the lender running appropriate queries in the data lake. That again, would not bode well for the transparency and timeliness, and hence, in the light of the argumentation above in the "customer's considerations" section, would not provide for an encouraging environment for vulnerabilities self-disclosures. Lastly, regarding the behavioral science aspects mentioned above, the data lake does not appear to contribute vastly to creating an environment stimulating self-disclosures, as there does not seem to be a major improvement in transparency with the lenders simply gaining another tool to manage their customers' data.

### 4.2 Blockchain technology evaluation

We have explored limitations of the data lake and to obtain a more holistic perspective would also need to evaluate blockchain for this use case.

In this case, specific customer data collection with regards to a particular product (such as a mortgage) could be managed on a single chain throughout the product's lifetime. Due to privacy concerns and relevant data protection laws, the transparency inherent within blockchain, which allows anyone to be able to inspect it, would need to be curtailed. That, however, is not a problem, as private blockchains – visible only to predefined parties – are already in use across several industries. In this case, a private blockchain could be utilized and designed in such a way that only the customer, the lender, and, if relevant, a mortgage broker, personal financial adviser/wealth manager, and, perhaps, the credit scoring agency could access the information held on the chain; with the ability to write further restrictions as necessary. Particularly with credit scoring agencies in mind, smart contracts representing events in the customer's history and stored on the blockchain could provide data-backed evidence on how this customer has handled their vulnerabilities before.

Interestingly, since the major CSPs – including Azure, AWS, and GCP – provide not just data lake but also blockchain solutions, banks, who are heavily invested in their data lakes, could potentially build on these with blockchain in a way that one technology could complement the other for the purpose of managing their customers' vulnerability self-disclosure data.

**Table 3:** Evaluation of the application of blockchain technology considering GDPR requirements

| GDPR REQUIREMENTS | EVALUATION OF THE APPLICATION OF BLOCKCHAIN TECHNOLOGY |
|---|---|
| Art. 13 GDPR "Information to be provided where personal data are collected from the data subject"[16] | Meets this article as smart contracts could be set up in a way that all the GDPR-required information would be provided to the data subject (i.e., customer flagging their vulnerability). |
| Art. 14 GDPR "Information to be provided where personal data have not been obtained from the data subject"[17] | In the event that vulnerability-related data stored on blockchain relating to the data subject were obtained via another party, the smart contract could be set up in a way that it would inform the customer of all the information required by GDPR. |
| Art. 15 GDPR "Right of access by the data subject"[18] | This is the area where the application of blockchain would have the clearest advantage over the application of data lake because it would offer the data subject the ability to instantly inspect data held on them by the data controller (i.e., the lender). |

[16] https://tinyurl.com/4xzbuutw
[17] https://tinyurl.com/a9ancxcu
[18] https://tinyurl.com/bdcmbmym

At first glance, the unrestricted transparency that comes with the use of public blockchain (in contrast to private blockchain proposed here) might make the application of this technology to managing self-disclosures of customers' vulnerabilities appear to go against the requirements of GDPR. As such, one may overlook the fact that it enables instant inspection of personal data held by the lender. Consequently, such a facility would in turn be very much GDPR-compliant. In fact, the use of blockchain would help the solution to meet GDPR articles as outlined in Table 3.

As such, it seems fair to say that blockchain does contribute to an increase in transparency and helps build an environment where customers should feel more comfortable to self-disclose their vulnerabilities. It also scores higher in terms of the aforementioned behavioral science criteria than a data lake solution.

**Table 4:** Evaluation of the application of blockchain technology considering Consumer Duty requirements

| FCA REQUIREMENTS | EVALUATION OF THE APPLICATION OF BLOCKCHAIN TECHNOLOGY |
|---|---|
| **1.16** which requires "firms to consider the needs, characteristics and objectives of their customers – including those with characteristics of vulnerability – and how they behave, at every stage of the customer journey." – raising the bar of consumer protection required of the regulated companies.<br><br>**8.5** which highlights that "many respondents queried the practical application (…) considerations relating to potential vulnerabilities; and the proposed approach to testing communications" – thereby proving the interest expressed by those consulted in the practicalities of flagging up potential vulnerabilities. | Evaluating the application of blockchain in the light of the relevant FCA requirements, as listed in the "Introduction", similar to the above data lake assessment, also here both the paragraphs 1.16 and 8.5 are met, as in either case, the new technologies help firms to better consider the characteristics of vulnerability of their customers and assist these customers with flagging up their vulnerabilities. However, blockchain could offer more with respect to communicating with the customer with regards to their vulnerabilities, as the fact that it allows all parties to write to it means it is more interactive than a data lake, which would be managed by the lender with inputs from other systems and the customer only allowed a limited insight. |
| **10.6** presents the view of the consumer organizations which "suggested that firms should be required to take an inclusive design approach to meet the needs of customers with characteristics of vulnerability" – hence clarifying the user requirement with regards to the functionalities for flagging potential and actual vulnerabilities. | Blockchain meets para. 10.6, as it allows the customer to write directly to the blockchain as well as to inspect in real time everything stored on it with regards to their data. It is more interactive and transparent, and, therefore, ticks the box of the "inclusive design approach", which the FCA specifically points to. |
| **Annex B 2A.7.4 G** specifies that "in relation to the needs and characteristics of retail customers, a firm should, among other things: (…)<br>**(4)** assist frontline staff to understand how to actively identify information that could indicate vulnerability and, where relevant, seek information from retail customers with characteristics of vulnerability that will allow staff to respond to their needs;" – thereby clarifying the requirement for the vulnerability flagging functionality from the point of view of the financial institutions, as well as | With regards to Annex B 2A.7.4 G (4), unlike in the case of the data lake, the fact that the vulnerable customers would use blockchain functionality to self-disclose and classify their problems means that it would assist frontline staff in identification of the information pertaining to these self-disclosed vulnerabilities. |
| "**(5)** set up systems and processes in a way that supports and enables retail customers with characteristics of vulnerability to disclose their needs;"[19] – hence providing a further clarification on the user requirement with respect to the same functionality. | Similarly, for 2A.7.4 G (5), also unlike the data lake, blockchain would help the customers self-disclose their vulnerabilities by allowing them to write directly to the chain. |

[19] https://tinyurl.com/yw6mm3p

## 5. CONCLUSION

In summary, the application of blockchain is the more appropriate solution to fulfill the FCA requirements with regards to vulnerability self-disclosures and complies with GDPR considerations. Blockchain enables the real-time capture of data directly from clients to create the data record at source (including future updates driven by changes in the client/consumer's personal circumstances). It provides the lender with the ability to proactively seek client information updates (through the application of smart contracts) as well as full auditability of the client/consumer data throughout the full product lifecycle and/or existence of the client relationship. It offers full and flexible control of the data through consensus and permissions by all participants in the chain (including the consumer). It fully supports consumer access to their data (in full compliance with GDPR) in a timely manner. Blockchain is also able to support the end-to-end client lifecycle management process through a single blockchain, removing the need to manage different stages of the process across multiple and disparate systems (leading to data integrity and quality issues).

Data lakes can also be considered as a valid solution and may have an advantage over blockchain as they are widely employed by financial services organizations today. However, the key disadvantage is that the data lake architecture tends to remove financial services organizations' proximity to client facing technology, which is required to capture client data and, therefore, makes it challenging to integrate valuable customer information with the same efficiency as blockchain.

Consequently, blockchain is the more transparent and inclusive option as it can allow the customer to write directly into it, enabling immediate inspection and, thereby, stimulating an honest, open dialogue between the parties.

Image generated by Adobe Firefly

As such, the behavioral science guidelines presented to empower the customers to self-disclose are also better fulfilled by blockchain. Today, the minimal penetration of traditional financial services by blockchain technology is a clear obstacle when it comes to adoption, as it may make this solution less cost effective than a data lake, even if the data lake does not meet all of the FCA's objectives with regards to the vulnerability self-disclosures set out in the Consumer Duty regulation. However, the analysis and assessment contained in this paper brings to light an innovative blockchain "use case" that financial services organizations should consider developing to facilitate and enhance their compliance with the Consumer Duty regulation.

# ASSESSING AI AND DATA PROTECTION EXPERTISE IN ACADEMIA AND THE FINANCIAL SERVICES SECTOR: INSIGHTS AND RECOMMENDATIONS FOR AI SKILLS DEVELOPMENT

**MARIA MOLONEY** | Senior Researcher and Consultant, PrivacyEngine, Adjunct Research Fellow, School of Computer Science, University College Dublin

**EKATERINA SVETLOVA** | Associate Professor, University of Twente

**CAL MUCKLEY** | Professor of Operational Risk in the Banking and Finance Area, UCD College of Business, and Fellow, UCD Geary Institute

**ELEFTHERIA G. PASCHALIDOU** | Ph.D. Candidate, School of Economics, Aristotle University of Thessaloniki

**IOANA COITA** | Consultant Researcher, Faculty of Economics, University of Oradea

**VALERIO POTI** | Professor of Finance, Business School, University College Dublin, and Director, UCD Smurfit Centre for Doctoral Research

## ABSTRACT

The proliferation of artificial intelligence (AI) is reshaping modern life in many ways. This has prompted action from many governments globally. The European Union is in the process of drafting a new E.U. AI Act, modeled on GDPR. To navigate this evolving regulatory landscape, financial researchers and industry professionals will need comprehensive training. However, existing efforts seem limited. This paper puts forth the idea of tailored training to better understand the complex interaction of data protection and ethical AI. It uses case studies to highlight the challenges of AI and the GDPR in the financial services sector. We also put forth survey findings that suggest current programs inadequately prepare individuals for GDPR compliance in AI. Recommendations include an initial training framework for ethical and compliant AI engagement.

## 1. INTRODUCTION

Artificial intelligence (AI) is revolutionizing contemporary life. The European Union (E.U.) acknowledges this progression of AI's central role in modern life and is crafting new regulation entitled the E.U. AI Act, which mirrors in many ways the influential General Data Protection Regulation (GDPR) – both are risk-based approaches to regulation. The financial sector is faced with an increasingly complex regulatory landscape and navigating it necessitates robust education. This paper argues that training efforts for this new landscape remain inadequate, particularly when faced with the existing GDPR, the forthcoming AI Act, and the expanding domain of AI ethics.

This paper underscores the need for targeted training through case studies, focusing on GDPR-compliant AI applications in the financial sector. A 2023 survey of financial experts and AI researchers investigates GDPR awareness and familiarity with Explainable AI (xAI). GDPR mandates transparency and accountability in AI decisions, achievable through xAI – an emerging field creating transparent AI models. GDPR-compliant xAI necessitates the fusion of not only GDPR principles into xAI development, but also interdisciplinary collaboration and transparent methods throughout the development lifecycle. The results of our survey indicate existing programs may not sufficiently prepare individuals in academia and industry for GDPR-compliant AI.

In conclusion, we offer an initial training framework to prepare academics and industry practitioners for ethical and compliant AI engagement, bridging the knowledge gap and ensuring GDPR compliance.

## 2. AI AND E.U. CITIZENS: THE WIDER PICTURE

Europe's approach to AI centers on promoting quality and trust while preserving human rights. Regulations like the E.U. Cybersecurity Strategy, Digital Services Act, Digital Markets Act, and Data Governance Act all contribute towards establishing a foundation for ensuring data quality. The Digital Europe Programme aims to bolster digital skills through education and training, addressing the gender gap, and fostering industry partnerships.

Convention 108+[1] and the GDPR are two of the leading canons for data protection in the digital age. To a degree, they both emphasize the importance of safeguarding personal data in AI applications. The AI Act, modeled on the GDPR, goes further in emphasizing data governance standards to prevent harmful biases in AI.

Concerns arise, however, regarding limited human oversight in AI-first business models that are particularly prevalent in the fintech sector. These business models aim to reduce costs by mitigating, or eliminating, human oversight and intervention. Recognizing this trend, the AI Act places stringent requirements on data used to train AI algorithms, emphasizing fairness and transparency, especially when personal data and special category data is used to train the algorithms.

The recent Dutch Child Benefit scandal exposed the potential harm of unchecked AI[2] use for financial matters. This case emphasizes the importance of balancing societal needs with the developmental advances of AI. GDPR compliance among fintech firms remains crucial. In fact, this supports the necessity for education and training. Horizon Europe, with a €95.5 billion budget, is prioritizing AI research and development, while aligning with the digital and green transition.

Questions linger regarding the extent to which the AI Act addresses prioritizing AI and data protection training across Europe and how rigorously the GDPR enforces transparency and accountability in AI technology. Further developments are needed to ensure responsible AI adoption and data protection.

### 2.1 GDPR and explainable AI (xAI)

Fintech companies are increasingly adopting AI technologies, with 75% expected to do so according to Gartner.[3] Ensuring GDPR compliance is vital in this context, as it requires organizations to provide a right to explanation when an automated decision is made about individuals. The European Parliament recognizes the GDPR's right to explanation as challenging for AI systems and believes that transparent AI decision-making processes could be achievable through the use of Explainable AI (xAI) methods.

The International Association of Privacy Professionals[4] asserts that xAI is pivotal for GDPR compliance. It aids individuals in understanding data usage and decision-making processes, reducing algorithmic bias risks in fintech. They emphasize the GDPR's right to explanation for building consumer trust and ethical AI use.

Financial research underscores xAI's role in enhancing transparency, decision accuracy, and customer trust in financial models. Challenges include data complexity, quality, and specialized knowledge.

Financial frameworks have in recent years advocated for GDPR-compliant xAI systems, emphasizing data protection and user rights. Compliance can mitigate legal liability and promote transparency, accountability, and data protection.

In conclusion, GDPR-compliant xAI is pivotal for responsible and ethical AI integration in fintech. It fosters consumer trust, minimizes risks, and drives AI-based innovation while ensuring GDPR adherence. As AI advances, embracing GDPR principles through xAI implementation remains essential.

### 2.2 Training on GDPR and AI Act in finance and academia: State-of-the-art

GDPR mandates significant changes in financial organizations regarding personal data management, encompassing secure archiving, transaction records, data processing, and customer rights. However, a Versasec survey[5] revealed that 27% of companies find educating internal employees challenging, and Deloitte noted a gap in data protection understanding

[1] https://tinyurl.com/3uxw476j
[2] https://tinyurl.com/342kaekj
[3] https://tinyurl.com/yaah3ykf
[4] https://tinyurl.com/y8nvr6r8
[5] https://tinyurl.com/yju8fwj4

between specialists and customer-facing staff.[6] Financial institutions have well-established training programs, adapted and expanded for GDPR, resulting in high compliance. Yet, AI specialists' GDPR training within finance remains unclear; for example, a German survey from the Center for Financial Studies in 2023 found that while 83% of respondents consider AI regulation necessary, details on the AI Act training are absent.[7]

In academia, Kenny (2021) found 71% of teachers and lecturers in Irish tertiary education participated in GDPR training, showing interest in additional training. However, GDPR awareness among academics in AI research remains underexplored, posing a research gap. Efforts to support GDPR education in academia should be considered, aligning with industry demands for AI-related GDPR compliance and ethical practices.

## 3. THE SURVEY

The prior discussion highlights the importance of financial experts' and academics' understanding of AI regulations and ethics. We conducted a survey in the winter of 2022/2023 to assess their current level of knowledge in this area.

### 3.1 Methodology

We conducted a survey with 89 participants, which consisted of AI users and developers in both academia and industry, particularly in fintech and finance. Respondents from 23 countries participated, mainly from, but not limited to, Europe. The average age of the respondents was 43, and 48% were female. The study employs various research designs, including comparisons between researchers and industry experts in AI expertise. We conducted a between-groups design to test the statistically significant difference in the perceived level of expertise in AI between researchers and industry experts. Building upon this result, we then sought to search for statistically significant associations. Due to sparse data and small sample sizes, we use Fisher's exact test based on N = 5e8 Monte Carlo simulations, which is employed when dealing with sparse data or small sample sizes to assess associations between variables, offering a more accurate assessment than the chi-square test in such cases.

### 3.2 Results

Results revealed a significant difference in perceived AI expertise between researchers and industry experts ($t = -3.6565$, $df = 60$, $p < 0.01$, two-tailed). Industry experts had higher perceived expertise levels (mean = 77.92, SD = 20.45) than researchers (mean = 49.52, SD = 24.92), supporting the idea that academics' perceived level of expertise in AI has an association with their level of knowledge of explainable AI (xAI) and their level of engagement with AI.

There was a significant relationship between knowledge of xAI and frequency of research into AI ($p = 0.0738$). Those conducting AI research as their primary area were more likely to know xAI principles (Table 1). xAI principles, emphasizing transparency and interpretability, equip academics to navigate AI complexities effectively. This understanding enables them to scrutinize AI ethics, biases, and societal impacts, vital aspects of scholarly AI inquiry.

Table 2 is a frequency table that shows responses in percentages as to whether respondents or their employers have ever suffered a data breach. The respondents are categorized according to whether they believe they have knowledge of the GDPR legislation or not.

Remarkably, a majority of respondents, regardless of their GDPR knowledge, claimed that they had never experienced data breaches, even those who considered themselves well-versed in GDPR. This pattern extended to their employers, with 60% of researchers and 89% of industry experts,

**Table 1:** Cross-tabulation of knowledge of xAI principles by the research conducted

| | | | | | | |
|---|---|---|---|---|---|---|
| KNOWLEDGE OF XAI | No | 2 | 1 | 5 | 6 | 2 |
| | Yes | 7 | 10 | 5 | 12 | 0 |
| | | Secondary research area | Main research area | Rarely | Sometimes | Never |

[6] https://tinyurl.com/y9fefks5
[7] https://tinyurl.com/ycany3m3

**Table 2:** Frequency table of data breaches among respondents with and without knowledge of the GDPR legislation

| | | ACADEMIA | | INDUSTRY | |
|---|---|---|---|---|---|
| | | Yes | No | Yes | No |
| **Knowledge of the GDPR legislation** | Personal data breach victim | 40% | 60% | 20% | 80% |
| | Employer is a data breach victim | 40% | 60% | 11% | 89% |
| **No knowledge of the GDPR legislation** | Personal data breach victim | 29% | 71% | 0% | 100% |
| | Employer is a data breach victim | 0% | 100% | 0% | 100% |

knowledgeable of the GDPR legislation, believing their organizations had never experienced a data breach.

Curiously, those who admitted to lacking GDPR knowledge reported no personal (71% of researchers and 100% of industry experts) or employer-related data breaches (100% for both). This contrasts with recent findings, such as the IDC cloud security survey revealing that 98% of companies faced cloud data breaches in the past 18 months[8] and the Egress' Insider Data Breach Survey 2021, which disclosed that 94% of organizations experienced insider data breaches.[9]

These discrepancies underscore the need for enhanced training and awareness of data security measures. Bridging this knowledge gap through targeted programs in academia and industry is crucial. Such initiatives can equip individuals and organizations with the skills to protect sensitive data effectively. Inadequate GDPR understanding may lead to inadequate, yet vital data protection measures, increasing the risk of data breaches and regulatory repercussions. Comprehensively addressing this gap is essential to fortifying data security, benefiting both academic and industry stakeholders.

Table 3 summarizes our survey results, emphasizing the relationship between training, expertise, and data protection practices among academics and industry professionals.

The survey revealed a significant difference in perceived AI expertise between academics and industry professionals, with the latter showing higher perceived expertise. Surprisingly, most in both groups who admitted to lacking GDPR knowledge received general data protection training (71% in academia, 75% in industry) but lacked data protection training specifically for AI (ranging from 0% to 25%). This is concerning since both groups develop AI models in their roles (33% for both according to Figure 1).

The difference in AI expertise between academics and

**Table 3:** Frequency table of received form of training among respondents with and without knowledge of the GDPR legislation

| | | ACADEMIA | INDUSTRY |
|---|---|---|---|
| **With knowledge of the GDPR legislation** | GDPR general training | 37% | 54% |
| | GDPR training for AI use | 2% | 6% |
| | GDPR training for AI development | 0% | 0% |
| | GDPR training for fintech | 2% | 14% |
| | No GDPR training | 58% | 43% |
| **Without knowledge of the GDPR legislation** | Data protection general training | 71% | 75% |
| | Data protection training for AI use | 0% | 25% |
| | Data protection training for AI development | 0% | 0% |
| | Data protection training for fintech | 43% | 25% |

Note: In certain questions, respondents were permitted to select multiple responses, allowing for the possibility of more than one option being chosen for each question.

8 https://tinyurl.com/3uw6smhm
9 https://tinyurl.com/fnz8ysyu

industry professionals may stem from a lack of specialized GDPR/AI training programs. Academics often focus on niche research areas, while industry experts require broader AI knowledge. Our findings also highlight the association between researchers knowledgeable about xAI principles and their primary focus on AI research. This underscores the importance of specialized education in cutting-edge AI concepts. Currently, AI developers lack awareness of Privacy by Design and data protection, as revealed by the survey. The AI Act mandates training to minimize erroneous or biased AI-assisted decisions in critical domains. While universities are beginning to offer data protection postgraduate education, the integration of AI-specific training, particularly emphasizing data protection and ethics, remains crucial for safeguarding individual rights under the GDPR.

The next section outlines two use cases where training, with an emphasis on data protection and ethics, in AI specific challenges, is required to ensure the rights and freedoms of individuals regarding their personal information under the GDPR are to be protected and maintained into the future.

## 4. TWO EXAMPLES TO ILLUSTRATE THE NEED FOR TRAINING IN THE FINANCE SECTOR

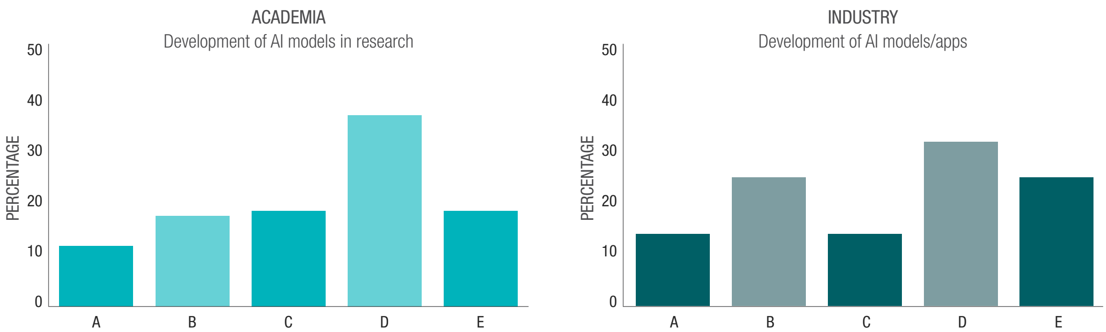
Let us look at two examples that illustrate the challenges of AI deployment in finance: AI informed algorithmic lending and AI-based management of clients' wealth. These examples serve to clarify relevant training needs for professionals in the financial services and fintech industries.

### 4.1 Example 1: AI informed algorithmic lending systems

Training of financial services' practitioners in algorithmic lending should not only include evidence of the prevalence of the use case but also its perilous nature. Such training further needs to span guardrails that can inform ethical algorithmic lending. As training of financial services personnel is imperative in the context of algorithmic lending, we use it to illustrate a stylized training program. Such a training program should pertain to the prevalence, peril, and the potential of algorithmic lending.

- **Prevalence of AI-informed algorithmic lending:** Interest in AI-informed algorithmic lending is widespread in financial services. It is highlighted, for instance, as a prevalent use case in surveys of financial services practitioners[10] and in industry[11] and regulator position papers.[12]
- **Peril of AI-informed algorithmic lending:** In the E.U.'s AI Act, algorithmic lending falls under the high-risk category in financial services. This designation stems from its potential to perpetuate and worsen unjust discrimination, leading to increased disparities in income and social status. This discrimination occurs when lending practices disproportionately affect minority "special category" borrowers, regardless of their creditworthiness. It can happen due to inadequate representation of minority borrower data, inclusion of variables correlated with sensitive minority classes (such as applicant's geographical location or job profile), and biased historical

**Figure 1:** Distribution of respondents developing AI models




A – Yes, developing AI models is my main area of research
B – Yes, I often develop AI models in my research
C – Yes, I sometimes develop AI models in my research
D – I rarely develop AI models in my research
E – No, I never develop AI models in my research

A – Yes, it is central to my work
B – Yes, it is not central to my work, but is important to my work
C – Yes, it is central to my company, but I do not work directly in it
D – Yes, it is important to my company, but I do not work directly in it
E – No, it is not important to me or my company

[10] https://tinyurl.com/yfhcxr87
[11] https://tinyurl.com/7exw686w
[12] https://tinyurl.com/yh2xta3c

loan officer decisions present in the training data. AI-driven lending, without properly trained financial personnel, is likely to result in unacceptable discrimination. This can severely impact individuals' rights and freedoms, making it crucial to address these issues in regulatory frameworks.

- **Potential of AI-informed ethical algorithmic lending:** Training is required in how to record, safeguard, and use sensitive data in the pre-, in-, and post-processing stages of algorithmic lending.

### 4.2 Example 2: AI-based decision making in the management of client's money

AI algorithms have been increasingly used in the investment industry to enhance the investment process and attract clients. In 2017, Tyndaris SAM, a hedge fund based in Monaco, entered into an agreement with the famous tycoon Samathur Li Kin-kan (represented by an investment vehicle VWM Limited) to manage his account using an artificial intelligence system called the K1 supercomputer. Li was promised by Tyndaris' CEO that this AI system had the capability to apply machine learning to process social media data and real-time news to predict investor sentiment in the financial markets and, based on this, the movements of U.S. stock futures; the program would then independently send instructions to a broker to execute trades, adjusting its strategy over time based on what it had learned. Although the strategy had been allegedly extensively back-tested and live-tested before trading began, VWM experienced substantial losses, including a U.S.$20 million loss on February 14, 2018. VWM terminated the contract and sued Tyndaris and its CEO for misrepresenting the K1 supercomputer's capabilities.[13]

Leaving legal issues aside, the case highlights the necessity for training of the involved employees at Tyndaris. First of all, investment professionals should be able to properly classify algorithms in use into one of four risk categories in accordance with the E.U. AI Act. Secondly, as the described situation implies the "human-on-the-loop" constellation (when AI users do not intervene into decisions of algorithms), responsible asset managers should be trained in how to efficiently exercise human oversight in such contexts and stay compliant with the regulation. Are there more specific requirements for data collection, data cleaning, and programming? Or, are more specific measures required for assessing outputs of AI decisions and trades? Thirdly, possible approaches to solving the problem of accountability and responsibility might also be a part of the training. Finally, sales representatives should be specifically trained to adequately communicate AI capabilities with respect to investments as well as its limitations to clients. They should be able to explain the workings of the algorithm at a very general level in a manner accessible to the lay investor. This also requires specific training.

## 5. STEPS TOWARDS BETTER TRAINING IN GDPR, THE AI ACT, AND AI ETHICS

Having analyzed and discussed the survey evidence and the illustrative case studies, we now turn to provide a detailed overview of possible steps towards improved training in GDPR, the AI Act, and AI ethics.

A valuable document of which we avail to flesh out such training steps is a report on "AI ethics, training and awareness raising" from the E.U. SIENNA Project.[14] The report differentiates between training in academia and industry. It particularly highlights the value of education in AI issues based on paradigmatic, real-world cases that have raised ethical concerns with respect to AI in society. In part inspired by this recommendation, we have presented two such cases above (algorithmic lending and AI in wealth management). As exemplified by the cases we presented, the inclusion of a case-based component in training programs is important because the cases help to illustrate the importance – and elusiveness – of many AI-related ethical issues (such as privacy, transparency, explainability, accountability, etc.) and the tensions that inevitably arise when professionals try to implement ethical principles in practice. We will now sketch out several steps to inform training programs for finance professionals regarding our two use case examples in AI in financial services, and, more generally, for students and researchers in an educational setting.

### 5.1 Training steps in algorithmic lending

In the case of algorithmic lending, the focus of training might be on data. Financial services personnel need to be informed about specific obligations for the processing of special category data (e.g., on race, religion, or sexual orientation). It is a difficult area due to a conflict between the right to privacy of individuals and the need to mitigate impermissible discrimination, which requires that special category data of individuals are disclosed and analyzed. In the same vein, a non-discrimination audit of AI-informed lending can prove challenging due to a tension between the rights of individuals

[13] https://tinyurl.com/c4hahrtk
[14] https://tinyurl.com/38u7mhe2

to both privacy (e.g., Article 7 of the Charter of Fundamental Rights of the European Union 2000; data protection law) and non-discrimination (e.g., Article 21 of the Charter of Fundamental Rights of the European Union 2000 and the Racial Equality Directive).

Indeed, the legal basis for permitting financial institutions to process special category data subject to specified obligations under the GDPR, can be argued, is unclear. Van Bekkum and Zuiderveen Borgesius[15] elaborate, for instance, that GDPR can hinder the collection or use of sensitive personal data to mitigate discrimination while other research papers conclude to the contrary that non-discrimination audits can be compliant. In time, safeguards for a lawful non-discrimination audit are likely to be informed in the finalized AI Act and by case decisions in national and European courts. Meanwhile, some dedicated training sessions could thematize this tension using case studies.

A training program, in the algorithmic lending setting, can differentiate between pre-, in- and post-processing of data. At the stage of data pre-processing before algorithmic model fitting, training is potentially required in the transforming of input data to achieve fairness. The probability of a minority borrower who repays a loan can be increased and a fair training dataset can be resampled. An algorithmic lending model can then be trained on the resampled data, which has ascribed less, or no, sensitivity to the minority class of borrower.

In-processing of data during algorithmic model fitting would require training on how to add a fairness-oriented regularization term to the model to penalize the model for impermissible discrimination. In case of post-processing of data after algorithmic model fitting, training is potentially required in how best to reassign labels (loan granted/ declined) to applicants where there is most uncertainty in the model prediction. This should be implemented to the advantage of the minority borrowing class, to achieve fairness in lending.

## 5.2 Training steps in investment management

The Tyndaris case underscores the urgency of training investment professionals in compliance with the E.U. AI Act. Proper classification of algorithms into risk categories is crucial, demanding training to understand this process and its implications. In situations like "human-on-the-loop", responsible managers need training in effective human oversight to comply with regulations. Specific training is essential for data collection, cleaning, programming, and assessing AI outputs. Asset managers must communicate AI capabilities transparently, utilizing xAI to cater to diverse investor needs. Training should address accountability, transparency levels, and policies, especially for sales representatives explaining algorithms to clients. Specialized training is vital for ensuring data integrity, privacy, and adherence to ethical frameworks like the CFA Institute's guidelines. These training programs should cover topics ranging from risk classification and human oversight to transparent communication and data privacy, aligning with regulatory requirements and ethical standards.

## 5.3 Training steps in an academic setting

Academic training for researchers and students must cover data protection, accountability, transparency, and interpretability. Utilizing explainable AI (xAI) in education enhances comprehension of financial concepts and decision-making processes. These findings indicate that xAI significantly improves learning outcomes in fintech education, promoting transparency and understanding in complex financial contexts. By incorporating xAI, academia prepares future financial professionals to adeptly navigate data-driven challenges and utilize ethical AI systems, fostering a responsible and knowledgeable industry workforce.

[15] https://tinyurl.com/3ujcfmy9

## 6. CONCLUSION

The observed disparity in perceived AI expertise between academics and industry professionals, with academics appearing to exhibit a lower perceived level of expertise compared to their industry counterparts, may be attributed to several underlying factors. This phenomenon can be explained through the lens of specialization versus generalization, highlighting that academics often possess a deeper understanding of specific AI-related subjects, while industry experts tend to have a broader, more generalized comprehension spanning multiple domains. This discrepancy underscores the intricate dynamics within the AI ecosystem and the distinct roles played by academics and industry professionals in its advancement.

Firstly, academics are typically engaged in research and teaching roles within academic institutions. They dedicate substantial time and effort to delving deeply into specialized AI topics, contributing to the development of foundational knowledge and the exploration of cutting-edge research areas. This focused approach may result in a perception of lower overall AI expertise when compared to industry experts because their expertise is concentrated within narrower niches.

Conversely, industry professionals, in their diverse roles, often require a more generalized understanding of AI concepts. They apply AI technologies across various applications, necessitating a broader comprehension that spans multiple domains, including business, technology, ethics, and compliance. Their perceived higher level of expertise may stem from this adaptability and versatility in applying AI solutions to real-world challenges.

Furthermore, it is essential to consider the evolving landscape of AI research and development. While industry professionals may be seen as having a more immediate, practical understanding due to their hands-on experience, academics play a crucial role in advancing the field through foundational research and innovative ideas. The case studies and the described pre-, in-, and post- data processing training steps outlined in this work suggest that academia serves as a fertile ground for exploring and conceptualizing new AI training methodologies and strategies. This research contributes to shaping the future of AI, even though it may not always align with immediate industry practices.

In conclusion, the perceived difference in AI expertise between academics and industry professionals can be attributed to the specialized focus of academics and the broader, practical orientation of industry experts. Recognizing the complementary roles played by these two groups is pivotal in understanding the dynamics of AI knowledge dissemination and advancement. Research conducted within academia, while possibly contributing to the perception of lower overall expertise, nonetheless holds immense value in shaping the future of AI training and development.

**REFERENCES**

Kenny, A., 2021, "Data protection awareness survey of tertiary sector teachers and lecturers in Ireland," Technological University Dublin, https://tinyurl.com/y7s8aa63

## ABOUT CAPCO

Capco, a Wipro company, is a global technology and management consultancy focused in the financial services industry. Capco operates at the intersection of business and technology by combining innovative thinking with unrivalled industry knowledge to fast-track digital initiatives for banking and payments, capital markets, wealth and asset management, insurance, and the energy sector. Capco's cutting-edge ingenuity is brought to life through its award-winning Be Yourself At Work culture and diverse talent.

To learn more, visit www.capco.com or follow us on Facebook, YouTube, LinkedIn and Instagram.

## WORLDWIDE OFFICES

**APAC**
Bangalore – Electronic City
Bangalore – Sarjapur Road
Bangkok
Chennai
Dubai
Gurgaon
Hong Kong
Hyderabad
Kuala Lumpur
Mumbai
Pune
Singapore

**EUROPE**
Berlin
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Milan
Munich
Paris
Vienna
Warsaw
Zurich

**NORTH AMERICA**
Charlotte
Chicago
Dallas
Hartford
Houston
New York
Orlando
Toronto
Washington, DC

**SOUTH AMERICA**
Alphaville
São Paulo

**THE COVER IMAGE WAS CREATED USING JASPER AI, AN AI ART GENERATOR**



WWW.CAPCO.COM

CAPCO 25
a wipro company