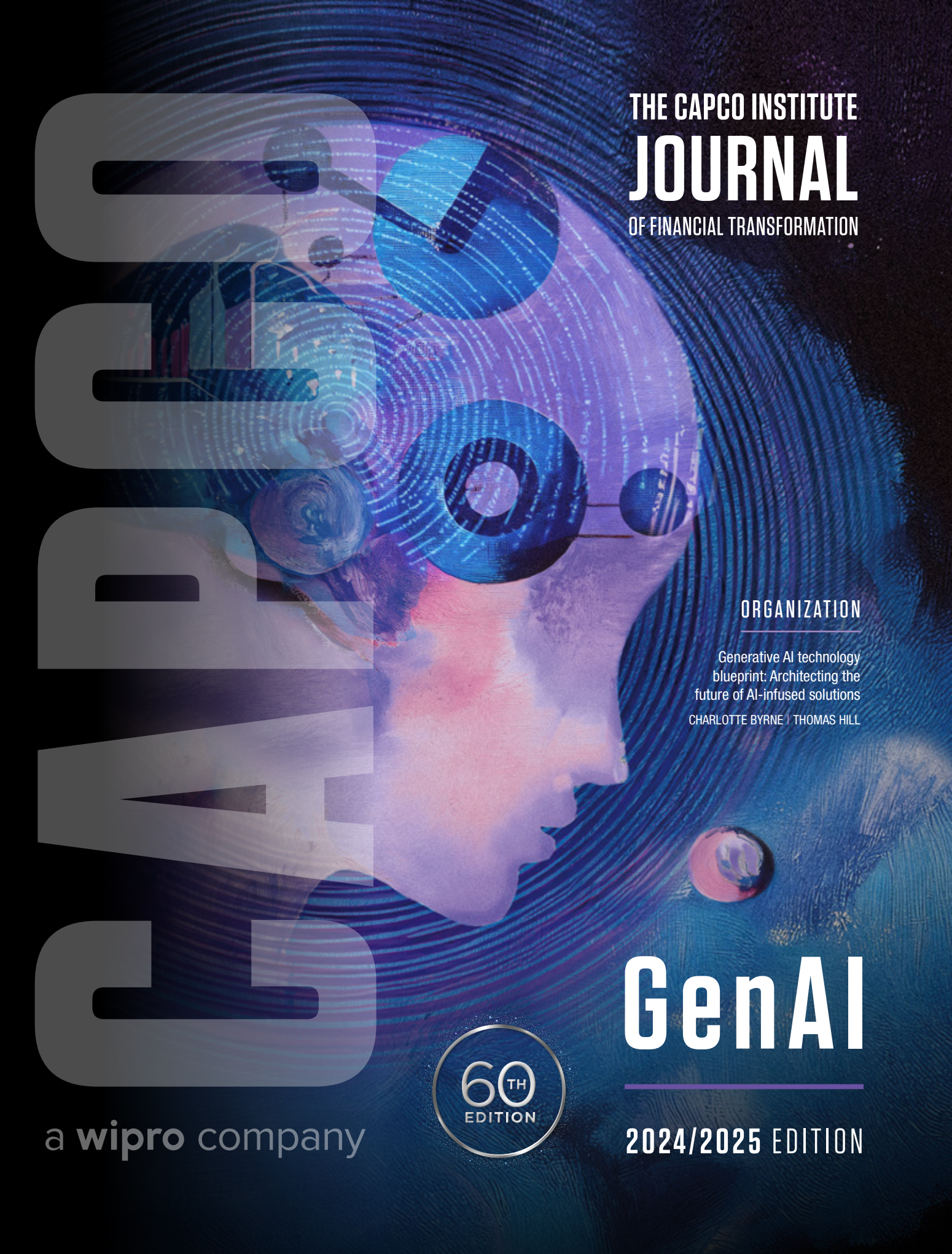


THE CAPCO INSTITUTE

JOURNAL

OF FINANCIAL TRANSFORMATION

ORGANIZATION

Generative AI technology blueprint: Architecting the future of AI-infused solutions

CHARLOTTE BYRNE | THOMAS HILL

GenAI

60TH EDITION

2024/2025 EDITION

a wipro company

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

Editor
Shahin Shojai, Global Head, Capco Institute

Advisory Board
Lance Levy, Strategic Advisor
Owen Jelf, Partner, Capco
Suzanne Muir, Partner, Capco
David Oxenstierna, Partner, Capco

Editorial Board
Franklin Allen, Professor of Finance and Economics and Executive Director of the Brevan Howard Centre, Imperial College London and Professor Emeritus of Finance and Economics, the Wharton School, University of Pennsylvania
Philippe d'Arvisenet, Advisor and former Group Chief Economist, BNP Paribas
Rudi Bogni, former Chief Executive Officer, UBS Private Banking
Bruno Bonati, Former Chairman of the Non-Executive Board, Zuger Kantonalbank, and President, Landis & Gyr Foundation
Dan Breznitz, Munk Chair of Innovation Studies, University of Toronto
Urs Birchler, Professor Emeritus of Banking, University of Zurich
Elena Carletti, Professor of Finance and Dean for Research, Bocconi University, Non-Executive Director, UniCredit S.p.A.
Lara Cathcart, Associate Professor of Finance, Imperial College Business School
Géry Daeninck, former CEO, Robeco
Jean Dermine, Professor of Banking and Finance, INSEAD
Douglas W. Diamond, Merton H. Miller Distinguished Service Professor of Finance, University of Chicago
Elroy Dimson, Emeritus Professor of Finance, London Business School
Nicholas Economides, Professor of Economics, New York University
Michael Enthoven, Chairman, NL Financial Investments
José Luis Escrivá, President, The Independent Authority for Fiscal Responsibility (AIReF), Spain
George Feiger, Pro-Vice-Chancellor and Executive Dean, Aston Business School
Gregorio de Felice, Head of Research and Chief Economist, Intesa Sanpaolo
Maribel Fernandez, Professor of Computer Science, King's College London
Allen Ferrell, Greenfield Professor of Securities Law, Harvard Law School
Peter Gomber, Full Professor, Chair of e-Finance, Goethe University Frankfurt
Wilfried Hauck, Managing Director, Statera Financial Management GmbH
Pierre Hillion, The de Picciotto Professor of Alternative Investments, INSEAD
Andrei A. Kirilenko, Reader in Finance, Cambridge Judge Business School, University of Cambridge
Katja Langenbucher, Professor of Banking and Corporate Law, House of Finance, Goethe University Frankfurt
Mitchel Lenson, Former Group Chief Information Officer, Deutsche Bank
David T. Llewellyn, Professor Emeritus of Money and Banking, Loughborough University
Eva Lomnicka, Professor of Law, Dickson Poon School of Law, King's College London
Donald A. Marchand, Professor Emeritus of Strategy and Information Management, IMD
Colin Mayer, Peter Moores Professor of Management Studies, Oxford University
Francesca Medda, Professor of Applied Economics and Finance, and Director of UCL Institute of Finance & Technology, University College London
Pierpaolo Montana, Group Chief Risk Officer, Mediobanca
John Taysom, Visiting Professor of Computer Science, UCL
D. Sykes Wilford, W. Frank Hipp Distinguished Chair in Business, The Citadel

CONTENTS

TECHNOLOGY

ORGANIZATION

REGULATION

CAPCO CEO WELCOME

DEAR READER,

Welcome to our very special 60th edition of the Capco Journal of Financial Transformation.

The release of this milestone edition, focused on GenAI, reinforces Capco's enduring role in leading conversations at the cutting edge of innovation, and driving the trends shaping the financial services sector.

There is no doubt that GenAI is revolutionizing industries and rapidly accelerating innovation, with the potential to fundamentally reshape how we identify and capitalize on opportunities for transformation.

At Capco, we are embracing an AI infused future today, leveraging the power of GenAI to increase efficiency, innovation and speed to market while ensuring that this technology is used in a pragmatic, secure, and responsible way.

In this edition of the Capco Journal, we are excited to share the expert insights of distinguished contributors across academia and the financial services industry, in addition to drawing on the practical experiences from Capco's industry, consulting, and technology SMEs.

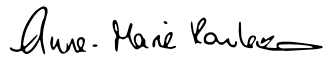
The authors in this edition offer fresh perspectives on the mindful use of GenAI and the implications of advanced GenAI on financial markets, in addition to providing practical and safe frameworks for boards and firms on how to approach GenAI governance.

The latest advancements in this rapidly evolving space demonstrate that the potential of GenAI goes beyond automating and augmenting tasks, to truly helping organizations redefine their business models, processes and workforce strategies. To unlock these benefits of GenAI, I believe that firms need a culture that encourages responsible experimentation and continuous learning across their organization, while assessing the impact of the potential benefits against a strategic approach and GenAI framework.

I am proud that Capco today remains committed to our culture of entrepreneurialism and innovation, harnessed in the foundation of our domain expertise across our global teams. I am proud that we remain committed to our mission to actively push boundaries, championing the ideas that are shaping the future of our industry, and making a genuine difference for our clients and customers – all while ensuring to lead with a strategy that puts sustained growth, integrity and security at the forefront of what we do.

I hope you'll find the articles in this edition both thought-provoking and valuable as you create your organization's GenAI strategy and future direction. As we navigate this journey together, now is the time to be bold, think big, and explore the possibilities.

My greatest thanks and appreciation to our contributors, readers, clients, and teams.



Annie Rowland, **Capco CEO**

GENERATIVE AI TECHNOLOGY BLUEPRINT: ARCHITECTING THE FUTURE OF AI-INFUSED SOLUTIONS

CHARLOTTE BYRNE | Managing Principal, Capco
THOMAS HILL | Principal Consultant, Capco

ABSTRACT

The generative AI (GenAI) landscape is evolving rapidly – and transforming how organizations approach and embrace technology and innovation. As businesses seek to harness the power of GenAI, it is crucial they establish a robust technology blueprint that guides the development, deployment, and management of AI-driven solutions. We explore the essential elements of a GenAI technology blueprint, covering the importance of flexible architectures, ethical considerations, and seamless integration with existing systems.

1. WHY A GENERATIVE AI TECHNOLOGY BLUEPRINT MATTERS

To effectively develop a GenAI technology blueprint, it is essential to recognize that GenAI is not the only factor shaping the future of technology – GenAI's synergy with the broader tech stack (including other artificial intelligence and machine learning tools), as well as the strength of an organization's data foundations, the robustness of past integrations, and the scope of cloud computing capabilities, will all have a profound impact.

A well-defined GenAI technology blueprint will serve as an invaluable roadmap, providing a structured approach to designing and implementing GenAI solutions that align with business objectives, while also addressing the unique challenges posed by GenAI.

By establishing clear architectural principles, governance frameworks, and integration strategies in advance, organizations can ensure the scalability, maintainability, and ethical deployment of GenAI solutions.

2. UNIQUE CHALLENGES POSED BY GENERATIVE AI

GenAI presents a set of distinct challenges that organizations must navigate to ensure successful adoption and deployment, some of which we highlight below.

Data quality and bias. GenAI models rely heavily on the quality and diversity of training data. Ensuring that the data used for training is representative, unbiased, and ethically sourced is a significant challenge. Biased data can lead to discriminatory or unfair outcomes, perpetuating societal biases in AI-generated content.

Intellectual property and content ownership. GenAI models have the ability to generate novel content, such as text, images, and audio. Determining the ownership and intellectual property rights associated with AI-generated content can be complex. Organizations must establish clear guidelines and legal frameworks to address issues related to content ownership, attribution, and licensing.

Explainability and interpretability. GenAI models, particularly deep learning-based models, can be highly complex and opaque. Understanding how these models arrive at their outputs and making their decision making processes interpretable is a significant challenge. Ensuring transparency and explainability is crucial for building trust in GenAI systems and meeting regulatory requirements.

Ethical considerations. GenAI raises ethical concerns related to privacy, fairness, and responsible use. Organizations must grapple with questions such as data privacy, consent, and the potential misuse of GenAI technologies for malicious purposes. Developing ethical frameworks and guidelines is essential to ensure the responsible deployment of GenAI solutions.

Integration with legacy systems. Integrating GenAI solutions with existing legacy systems can be challenging. Organizations must navigate compatibility issues, data integration challenges, and the need for seamless interoperability between GenAI components and traditional software systems. Overcoming these integration hurdles requires careful planning and robust integration strategies.

Talent and skills gap. The rapid advancement of GenAI technologies has created a talent and skills gap. Organizations face the challenge of acquiring and retaining employees with expertise in GenAI techniques, such as deep learning, natural language processing, and computer vision. Building internal capabilities and upskilling the existing workforce are crucial for successful GenAI adoption.

3. KEY COMPONENTS OF A GENERATIVE AI TECHNOLOGY BLUEPRINT

3.1 GenAI application architecture

A GenAI technology blueprint should outline a flexible and scalable application architecture designed to leverage the capabilities of generative models. The architecture should facilitate the creation of new, unique content using enterprise data and integrate seamlessly with current systems for diverse applications.

The following key components should be considered when establishing an application architecture:

- **Experience layer:** this layer encompasses various user interfaces, such as chatbots, contact center portals, web applications, and API playgrounds, enabling seamless interaction with GenAI solutions.

> *"Capco realized additional efficiencies of up to 50% in certain tasks where individuals had the right training."*

- **API management:** robust API management is crucial for facilitating integration between GenAI applications and external systems, ensuring secure and efficient data exchange.
- **GenAI platform:** the GenAI platform serves as the core of the architecture, providing orchestration and model management capabilities. It includes components such as prompt libraries, GenAI models (custom, open-source, and closed-source), and MLOps platforms for model training and deployment.
- **Data storage:** efficient data storage mechanisms, such as knowledge graphs, relational databases, data lakes, and vector databases, are essential for storing and retrieving relevant data for GenAI models.
- **Observability and monitoring:** comprehensive observability and monitoring tools are necessary to track the performance, usage, and outcomes of GenAI solutions, enabling continuous improvement and auditing.

3.2 Types of GenAI models and adaptation strategies

The technology blueprint should consider the various types of GenAI models available and provide guidance on adapting them to specific use cases. The blueprint should cover the following aspects:

- **Model catalog:** maintaining a comprehensive model catalog is crucial for managing and updating information about existing GenAI models, as well as integrating new models as they become available. The catalog should include details such as model types, use cases, performance benchmarks, architectures, and data requirements.

- **Model customization:** the blueprint should outline strategies for customizing GenAI models to specialize in specific domains or tasks. Techniques such as fine-tuning, adapter tuning, and reinforcement learning from human feedback (RLHF) can be employed to enhance model performance and adapt to specific requirements.
- **Retrieval augmented generation (RAG):** RAG is a powerful technique that combines retrieval mechanisms with generative models to provide more accurate and contextually relevant responses. The blueprint should provide guidance on implementing techniques like RAG, including data retrieval strategies, embedding techniques, and integration with GenAI models.
- **Prompt engineering:** effective prompt engineering is crucial for guiding GenAI models to generate desired outputs. The blueprint should cover best practices for crafting prompts, including techniques such as zero-shot, one-shot, and few-shot learning, as well as chain-of-thought prompting and prompt chaining.

3.3 Solution designs and rationale

The GenAI technology blueprint should provide standardized architectural designs and recommendations on how the architectural patterns have been applied to trending GenAI capabilities or similar sets of requirements. It should include:

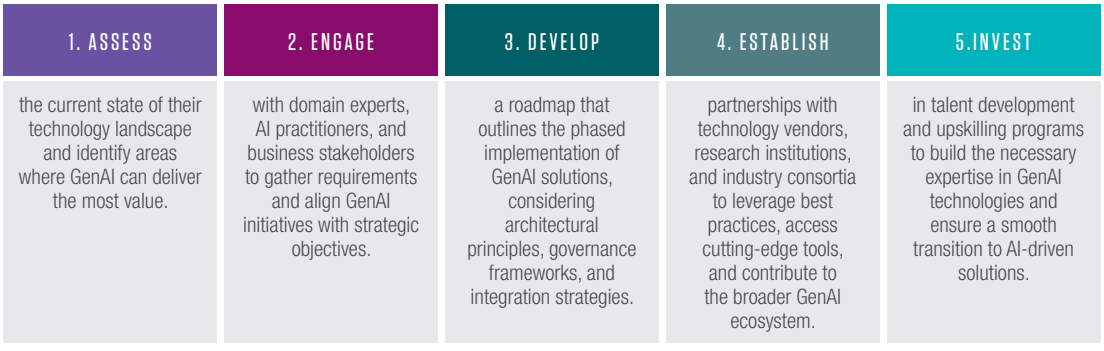
- **Data architecture:** the blueprint should outline the key data components and considerations for GenAI solutions, such as data discovery, profiling, sourcing, ownership, quality, metadata, and storage.
- **Technology stack:** the blueprint should recommend a suitable technology stack for implementing GenAI solutions, leveraging tools and services from leading cloud platforms such as Microsoft Azure, Amazon Web Services or Google Cloud Platform.
- **Deployment patterns:** the blueprint should provide guidance on deploying GenAI solutions using various patterns, such as containerization, serverless computing, and edge deployment, based on specific requirements and constraints.

3.4 GenAI LLM ops framework

The GenAI technology blueprint should include a framework for building and optimizing Large Language Model Operations (LLM Ops). The LLM Ops framework should cover the following aspects:

- **Model development:** guidelines for selecting the appropriate foundation models, training datasets, and architectures for GenAI model development.
- **Model deployment:** best practices for deploying GenAI models, including considerations for scalability, performance optimization, and monitoring.
- **Model maintenance:** strategies for maintaining and updating GenAI models, including version control, continuous integration and deployment (CI/CD) pipelines, and performance monitoring.
- **Governance and security:** frameworks for ensuring the ethical use, misuse prevention, and adherence to compliance standards in GenAI model development and deployment.

Figure 1: To embark on the journey of creating a GenAI technology blueprint, organizations should consider the following five steps.



4. CONCLUSION: GETTING STARTED WITH A GENERATIVE AI TECHNOLOGY BLUEPRINT

A well-crafted GenAI technology blueprint is a vital tool for organizations seeking to harness the transformative power of generative AI. By prioritizing flexible architectures, ethical considerations, seamless integration and continuous monitoring, organizations can accelerate their GenAI adoption and unlock new opportunities for innovation and growth.

As the GenAI landscape continues to evolve, organizations that invest in robust technology blueprints will be ideally positioned to navigate the challenges and opportunities ahead – effectively leveraging GenAI to drive transformative outcomes and shape the future of their industries.

© 2024 The Capital Markets Company (UK) Limited. All rights reserved.

This document was produced for information purposes only and is for the exclusive use of the recipient.

This publication has been prepared for general guidance purposes, and is indicative and subject to change. It does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (whether express or implied) is given as to the accuracy or completeness of the information contained in this publication and The Capital Markets Company BVBA and its affiliated companies globally (collectively "Capco") does not, to the extent permissible by law, assume any liability or duty of care for any consequences of the acts or omissions of those relying on information contained in this publication, or for any decision taken based upon it.

ABOUT CAPCO

Capco, a Wipro company, is a global management and technology consultancy specializing in driving transformation in the energy and financial services industries. 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 LinkedIn, Instagram, Facebook, and YouTube.

WORLDWIDE OFFICES

APAC
Bengaluru – Electronic City
Bengaluru – Sarjapur Road
Bangkok
Chennai
Gurugram
Hong Kong
Hyderabad
Kuala Lumpur
Mumbai
Pune
Singapore

MIDDLE EAST
Dubai

EUROPE
Berlin
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
Glasgow
London
Milan
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Houston
New York
Orlando
Toronto

SOUTH AMERICA
São Paulo

THIS UNIQUE IMAGE WAS GENERATED USING MID-JOURNEY, STABLE DIFFUSION AND ADOBE FIREFLY

WWW.CAPCO.COM

CAPCO
a wipro company