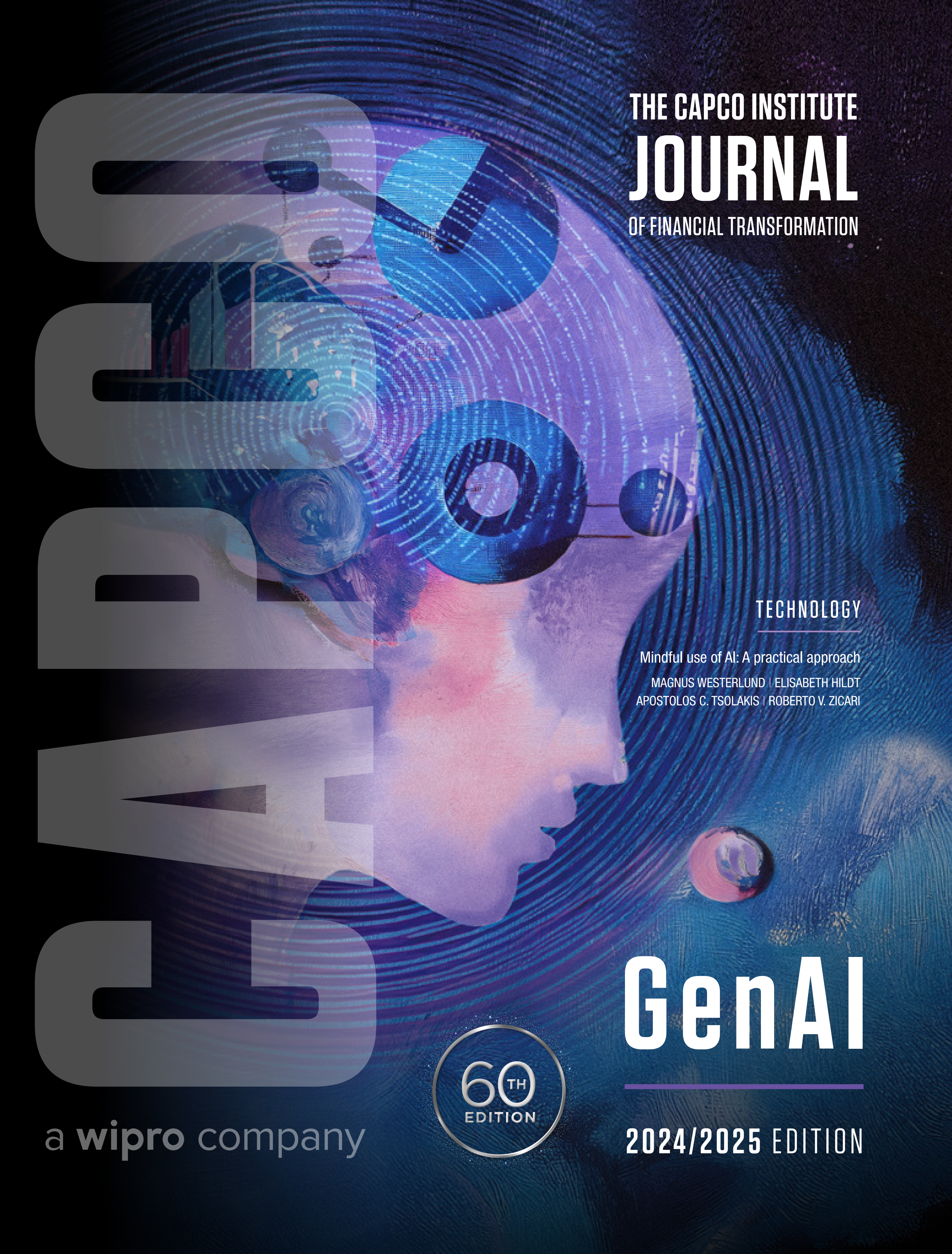


THE CAPCO INSTITUTE

JOURNAL

OF FINANCIAL TRANSFORMATION

TECHNOLOGY

Mindful use of AI: A practical approach

MAGNUS WESTERLUND | ELISABETH HILDT
APOSTOLOS C. TSOLAKIS | ROBERTO V. ZICARI

GenAI

2024/2025 EDITION

60TH EDITION

CAPCO

a wipro company

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

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**

MINDFUL USE OF AI: A PRACTICAL APPROACH[1]

MAGNUS WESTERLUND | Principal Lecturer in Information Technology and Director of the Laboratory for Trustworthy AI, Arcada University of Applied Sciences, Helsinki, Finland

ELISABETH HILDT | Affiliated Professor, Arcada University of Applied Sciences, Helsinki, Finland, and Professor of Philosophy and Director of the Center for the Study of Ethics in the Professions, Illinois Institute of Technology, Chicago, USA

APOSTOLOS C. TSOLAKIS | Senior Project Manager, Q-PLAN International Advisors PC, Thessaloniki, Greece

ROBERTO V. ZICARI | Affiliated Professor, Arcada University of Applied Sciences, Helsinki, Finland

ABSTRACT

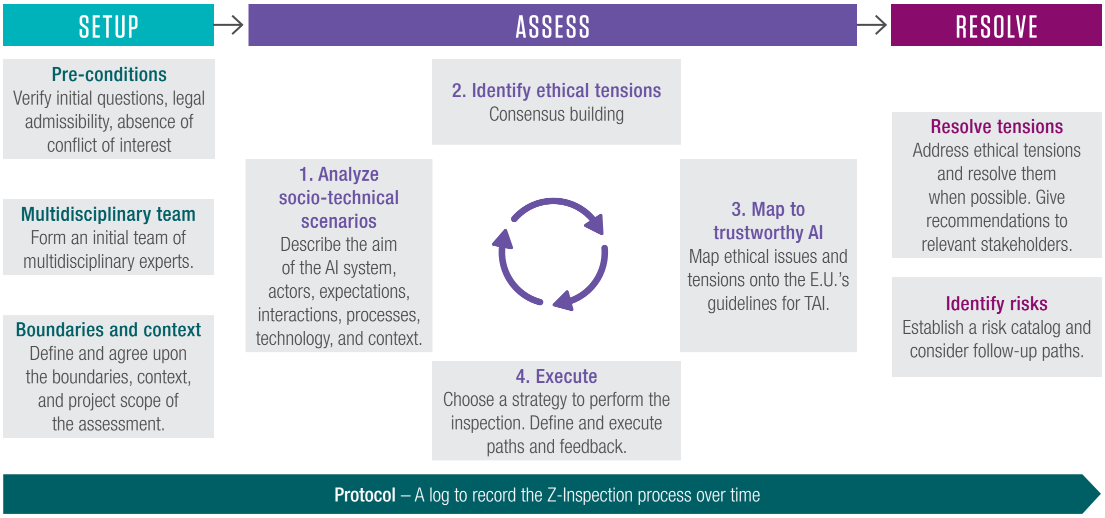
The current landscape of assuring AI reliability and quality is fragmented, with existing frameworks often lacking a unified methodology for comprehensive evaluation, particularly in integrating ethical and human rights considerations. This article introduces the Z-Inspection® process as a participatory, human-centered approach for assessing and co-designing trustworthy AI systems throughout their lifecycle. By forming multi-disciplinary teams and utilizing socio-technical scenarios, Z-Inspection® enables the exploration of ethical dilemmas and risks in context, fostering a shared understanding among stakeholders. This methodology aligns with the European AI Act's emphasis on human-centric technology and addresses limitations in existing standards by incorporating continuous ethical reflection and adaptability. We demonstrate how the co-design aspect of Z-Inspection® facilitates proactive risk identification, transparency, and alignment with regulatory requirements. This approach advances beyond traditional static checklists, offering a dynamic framework that intrinsically weaves ethical considerations into AI development, thereby ensuring that AI technologies are not only technically robust but also ethically sound, socially beneficial, aligned with human values, and legally compliant. Trustworthy AI is not an afterthought or technical hindrance but a way to promote a mindful use of AI.

1. INTRODUCTION

A fragmented approach to assessment and implementation characterizes the current landscape of assuring AI reliability and quality. Existing frameworks, such as the E.C. High-Level Expert Group on Artificial Intelligence (AI HLEG), National Institute of Standards and Technology (NIST), OECD, and Google's AI principles, provide valuable insights but lack a unified methodology for comprehensive evaluation. The work on standards (ISO, IEEE) and the CEN/CENELEC (cencenelec.eu) harmonized E.U. standards struggle to incorporate ethical and human rights aspects into the compliance and audit process. Particularly for high-risk AI systems in the public sector, conformity assessments are unlikely sufficient to determine the risk to human rights and ethics when considering AI systems such as generative AI (GenAI). Concept-based assessments focusing on individual aspects like accountability, fairness, and explainability have been developed, but they often operate in isolation, failing to capture the interdependencies of these elements. Industry-specific frameworks (e.g., IEEE P2247.4) and human rights-based approaches [Dutch Fundamental Rights and Algorithm Impact Assessment (FRAIA), Council of Europe Framework Convention on AI)] have emerged, but their integration into cohesive, widely applicable standards remains a challenge. Current explainable AI (XAI) solutions, while advancing rapidly, still struggle to balance robustness and efficiency with user-friendly interpretability, especially in complex

[1] This work was co-funded by the European Union under GA no. 101135782. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or CNECT. Neither the European Union nor the granting authority can be held responsible for them.

Figure 1: Z-Inspection® process for trustworthy AI



Z-Inspection® process flow describing the main steps of the setup, assess, and resolve phases. In parallel to the phases, a log is kept in which the process and events of the assessment are tracked. Adapted from Zicari et al. (2021b)[2]

domains like automated systems and GenAI. Moreover, the ethical implications and human rights considerations in AI development and deployment are often treated as secondary concerns, rather than being intrinsically woven into the fabric of AI systems from conception to implementation.

The Z-Inspection®[3] process for trustworthy AI (Figure 1) offers a different path to assessing AI trustworthiness throughout the AI system lifecycle.[4] Z-Inspection® is a validated participatory process based on human expertise that follows the AI HLEG requirements and breaks them down to deliver an ethical understanding of issues regarding specific AI use.[5] The Z-Inspection® process can be applied to the entire AI lifecycle, typically including (1) design, (2) development, (3) deployment, (4) monitoring, and (5) decommissioning.

Recent Z-Inspection® work includes a study in collaboration with the Dutch government to combine the trustworthy AI assessment with an FRAIA[6] that was accomplished with great success. The work highlights the importance of capturing future intentions early. It also emphasizes considering how people may be affected, by developing socio-technical scenarios that consider the broader contextual use of AI technology. This helps to avoid, for example, system-of-systems issues when model output propagates. Such issues are complex to capture with a product-centric regulation or standard and demand a broader discussion.

The same approach can also be employed to co-design trustworthy AI systems. The socio-technical scenarios can be developed early on, during the requirements elicitation, together with the technology providers implementing the AI system. Key insights can help define a more complete set of non-functional system requirements while also guiding the core functionalities of the system, i.e., the functional requirements, towards a system architecture that is more likely to deliver trustworthy results.

2. EUROPEAN AI ACT

The European Commission has introduced a regulation that wants to "ensure a consistent and high level of protection of public interests as regards health, safety and fundamental rights" (AI Act, recital 7). The ambition is that all deployed AI systems in the E.U, are based on human-centric technology, with the ultimate aim of AI increasing human wellbeing, especially considering the risk level of an AI system (Figure 2).

[2] Vetter, D., et al., 2023, "Lessons learned from assessing trustworthy AI in practice," Digital Society 2:3, 35
[3] Z-Inspection is a registered trademark distributed under the terms and conditions of the Creative Commons (Attribution-NonCommercial-ShareAlike CC BY-NC-SA) license (z-inspection.org)
[4] Zicari, R. V., et al., 2021, "Z-Inspection: a process to assess trustworthy AI," IEEE Transactions on Technology and Society 2:2, 83-97
[5] Allahabadi, H., et al., 2022, "Assessing trustworthy AI in times of COVID-19: deep learning for predicting a multiregional score conveying the degree of lung compromise in COVID-19 patients," IEEE Transactions on Technology and Society 3:4, 272-289
[6] Gerards, J., M. T. Schäfer, I. Muis, and A. Vankan, 2021, "Fundamental rights and algorithms impact assessment (FRAIA)," Rijksoverheid, https://tinyurl.com/y75hfh5s

The legislation is influenced by the definition of trustworthy AI (TAI), and by enacting the regulation, the Commission considers it a key aspect of Europe being a leader in TAI solutions. TAI was defined by the Commission's appointed High-Level Expert Group on Artificial Intelligence in 2019 and is based on four "ethical principles" – (1) respect for human autonomy, (2) prevention of harm, (3) fairness, and (4) explicability – and seven "requirements" that are closely related to these principles: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) societal and environmental wellbeing, and (7) accountability. The ethical principles are considered imperatives that AI practitioners should always adhere to. However, the HLEG already foresaw that the situation may arise where there are tensions between the principles and that new requirements will emerge as the technology develops and the use of AI becomes more integrated. The HLEG developed an initial checklist for practitioners to consider, but as the field has evolved, this checklist can no longer be considered complete. Furthermore, the limits of using predetermined checklists are that they are usually not dynamic enough to capture ethical reasoning.

3. HARMONIZED STANDARDS

By crafting harmonized E.U. standards that organizations can use to certify their solutions, the E.U. hopes to make the implementation of the AI Act easier than is the case with, for example, the General Data Protection Regulation (GDPR). The work for harmonized standards was given to the CEN/CENELEC standardization body and has yet to be completed. Harmonized standards will be created in collaboration with other international standardization bodies. However, there are some specific legal mandates that require new perspectives. One such standard is the conformity assessment standard, which should define the scope of what companies should deliver to ensure compliance.

Figure 2: The four-level risk-based approach defined within the AI Act

Adapted from: https://tinyurl.com/bcjsjkd9

The current preparation for a conformity assessment technical report has revealed that the requirements for ethical concerns are not directly part of what CEN/CENELEC can deliver. A fundamental difficulty in assessing ethical concerns for the purpose of a certification is that responses are not binary (pass/not pass) but may present dilemmas or a spectrum of voices that require further exploration. Treating the AI system as an isolated component makes it easier to audit technical conformity. However, the new E.U. legislation demands that the resulting AI system is trusted and trustworthy, and the problem lies herein. A company may receive a certification for a model, but integrating the model into a more extensive pipeline and the continuous operation of this system in a particular context is a very different problem than presented by the individual model. In fact, it has been shown by the MIT AI Risk Repository that most risks (65%) emerge after the AI system has been deployed.[7] Thus, we must ask, what is the value of certification if ethical or societal concerns are not addressed in the context of applying the AI, and if continuous use modifies the data, model, or pipeline, or, even more concerning, depends on a secondary model?

4. THE Z-INSPECTION® SELF-ASSESSMENT PROCESS FOR TRUSTWORTHY AI

Our work within the Z-Inspection® initiative has taken a different approach that aims to establish a self-assessment process for AI practitioners and procurement teams that want to evaluate an AI system/component in a real-world environment. The process is participatory and seeks to consider the AI HLEG principles and requirements by forming a representative multi-disciplinary team that covers each needed area. Following a structured approach, the assessment team develops an understanding of the use case, environment, and technology that allows them to project socio-technical scenarios. By using scenarios, the work allows for an exploration of past, present, and future considerations. The team then uses a meta-framework for the claim-arguments-evidence (CAE) analysis[8] of what was discovered to establish which claims are actual

[7] Slattery, P., et al., 2024, "The AI Risk Repository: a comprehensive meta-review, database, and taxonomy of risks from artificial intelligence," arXiv preprint arXiv:2408.12622

[8] Bloomfield, R., and J. Rushby, 2020, "Assurance 2.0: A manifesto," arXiv preprint arXiv:2004.10474

and which are not. Our assessment experience is that this is a very fruitful stage to establish a shared understanding of intentions and to limit the introduction of risks going forward.[9]

Following the establishment of actual claims, the process determines what evidence exists to support such claims. This work is usually done in a domain-specific manner by experts to allow for an in-depth study of concerns. Examples of such experts can be technical (machine learning and/or software architecture), legal, ethics, human rights, and, of course, domain experts from the actual environment where the system will be deployed, such as medical doctors, ecologists, and economists. Once the evidence has been gathered, it is shared among the entire group of experts, who can then still revisit their own conclusions. The final step is to verify the presented arguments that link the claims and evidence.

Based on the CAE review, an intermediary report is created and presented to the case owner, and the discussion is then aimed at resolving outstanding concerns. There are often situations that require ethics expertise, particularly to find and classify dilemmas as true or false. A vital aspect of the process is not to act as an authority that sits above the practitioners but rather as a council of peers that first helps define the solution, establish scenarios to reason within, and finally provide an outside view of what evidence is present that can validate the claims. Here, it is essential that the self-assessment team is constructed openly, without any competing interests or fear of retribution. Hence, optimally the developers themselves should not be part of the assembled expert team as they would have competing interests.

5. Z-INSPECTION® AS A TAI CO-DESIGN PROCESS

The Z-Inspection® methodology goes beyond many other TAI assessment frameworks by incorporating co-design principles throughout the AI system lifecycle.[10] This co-design approach is fundamental to addressing the complex, interdisciplinary challenges posed by AI systems, particularly in high-risk domains. Integrating human-centered TAI design principles into the development work ensures that the resulting framework is not only technically robust but also accessible and meaningful to end-users and practitioners.

By facilitating a co-design process, diverse stakeholders can work together from the early stages of AI system design. This interdisciplinary collaboration ensures that multiple perspectives are considered, from conception and requirements handling to development and systems testing. By utilizing a TAI co-design process, it leads to a more comprehensive understanding of potential impacts and risks, enabling the design of a more robust, reliable, and resilient system architecture.

5.1 Co-design use case example

The co-design aspect of Z-Inspection® promotes an iterative approach to AI system development. Rather than treating ethical and societal considerations as an afterthought or a one-time compliance check, the process encourages continuous evaluation and refinement throughout the AI lifecycle. An example of Z-Inspection® functioning as a co-design process can be found in the study "Co-design of a trustworthy AI system in healthcare: deep learning based skin lesion classifier."[11]

In this study, the co-design methodology was applied during the early design phase of an AI system intended to assist dermatologists in diagnosing skin lesions using deep learning algorithms. For the case study, dermatologists, evidence-based medicine experts, ethicists, and patient representatives were brought together with AI engineers. This diverse group identified ethical aspects and tensions between different viewpoints, such as the varying perspectives on overdiagnosis, early detection, and prognosis-based forecasting, which might have been overlooked in a traditional development process. This interdisciplinary input helped the group of researchers and practitioners developing the tool to shape and refine the design process.

5.2 Implementation of a TAI co-design process

The agile approach yields multiple advantages throughout the AI development lifecycle. It enables proactive risk identification, ensuring that potential issues are detected and mitigated before they become entrenched in the system architecture. The methodology facilitates real-time feedback on the AI system design, empowering development teams to iteratively enhance and refine the product based on continuous stakeholder involvement. Moreover, it fosters ongoing

[9] Vetter et al. (2023)

[10] Zicari, R. V., et al., 2021, "Co-design of a trustworthy AI system in healthcare: deep learning based skin lesion classifier," Frontiers in Human Dynamics 3, 688152

[11] Ibid.

alignment with social responsibility initiatives and evolving market expectations, ensuring the AI system maintains its relevance and ethical standing. Finally, this iterative framework cultivates organizational adaptability, reducing unforeseen AI reputational risks throughout the operational lifespan.

5.2.1 CO-DESIGN SETUP

The co-design approach starts with assembling a multi-disciplinary team comprising, for example, of AI engineers, domain experts, ethicists, legal experts, end-user representatives, and social scientists. The expert group works collaboratively to understand the AI system's aim, consider its potential impacts, and identify stakeholders' needs and concerns. Including various experts is crucial in i) understanding ethical, legal, and technical issues that could arise from the system's use, ii) assessing risks and harms, and iii) ensuring fairness.

An essential part of the setup is also to clearly define the scope of the project (including the boundaries and context of the assessment) and to create a detailed log of what is discussed and agreed to. This log will help to avoid scope creep, which often occurs in similar projects. This suggests that the team should understand the intended context and use of the AI system sufficiently to be able to, for example, analyze potential dual-use issues (unintended use of the AI).

5.2.2 SOCIO-TECHNICAL SCENARIOS

Similarly to the one-off assessment, the co-design approach uses socio-technical scenarios to establish a shared understanding of motivations and claims. Socio-technical scenarios involve the societal and technical context in which an AI system is (expected) to be used. This broad perspective avoids a narrow view in which only the tool itself and its technical aspects are assessed. These scenarios serve as a participatory design tool, enabling stakeholders to envision and explore various potential uses and impacts of the AI system in real-world contexts.

During the initial design phase of the AI system, we can start by defining TAI-related non-functional requirements and analyzing the technical functional requirements. In a current case study, an E.U. funded Horizon Europe project, "MANOLO" (GA 101135782),[12] we have employed this method to understand the AI components and system architecture that the project will deliver. In addition to requirements handling, as this is a project that starts with a low technology readiness level (TRL), we included comprehensive desk research to proactively identify potential dependencies and consequences that may later become concerns or dilemmas. Through this approach, we hope to bridge the gap between technical capabilities and practical applications.

5.2.3 IDENTIFYING ETHICAL ISSUES AND TENSIONS

To identify ethical issues and tensions when co-designing a trustworthy AI system, we convene the multi-disciplinary team of experts and thoroughly review the proposed AI system, its intended use, and potential impacts. Depending on the project phase, the information in terms of claims, arguments, and evidence will be more or less detailed. The experts use the information that is currently available to conduct one or more structured brainstorming session(s) to surface potential ethical, legal, and technical concerns from different perspectives, considering the impacts on various stakeholders. The outcome can then be categorized and prioritized, and the identified issues are delimited into common categories like privacy and data protection, fairness and non-discrimination, transparency and explicability, safety and robustness, human agency and oversight, and accountability. Here, the ethical principles and requirements delineated in the Ethics Guidelines for Trustworthy AI serve as guidance. In a process that serves to formalize the findings, the identified ethical issues are brought in line with and mapped to the ethical principles and requirements of the European guidelines document.

The second part of this step is to analyze potential tensions between different ethical principles or stakeholder needs, such as privacy versus model performance, or explainability versus accuracy. Each identified issue and tension are documented in detail, including rationale and potential implications. To validate findings, a consensus-seeking discussion takes over to ensure the correctness both in terms of the project (scope and intentions) and also as a consensus of experts regarding the detailed issues. Finally, mitigation strategies are sought for high-priority issues, considering technical, operational, and governance approaches.

Going forward, for each new development phase, we revisit and update the ethical analysis regularly as the design and stakeholder demands evolve. This structured approach allows for the comprehensive identification of ethical concerns from multiple perspectives early in the design process. It follows along with the implementation process to ensure that the initial

[12] https://tinyurl.com/rrvhmhr8

envisioned outcome is delivered together with the necessary evidence to support the final assessment. In fact, due to the trustworthiness related requirements and the evidence logged along the way, the final assessment becomes much more structured, facilitating the overall process of reaching a conclusion of whether the AI system assessed is trustworthy or not.

5.2.4 PRODUCTION

When a co-design product is launched into production, we enter into a new phase of the AI lifecycle that is often outside the scope of the co-design development project. However, the alignment work during the development phases of the project means that maintenance and governance practitioners have a strong body of support, and the stochastic and non-deterministic nature of the AI system may cause fewer surprises.[13] Hence, a core outcome of the co-design approach is that it fosters transparency in the AI development process while establishing pipelines that can generate evidence toward a long-term trustworthy outcome. Involving diverse stakeholders and encouraging open dialogue helps build trust among all parties involved, including potential end-users and the broader public.

5.2.5 RISK IDENTIFICATION, MONITORING, AND COMPLIANCE

During a complex project, it may sometimes be difficult or even impossible to mitigate and resolve every identified ethical issue. The co-design approach can also be extended to include a step that identifies risks that need to be monitored and tracked throughout the design and while in production. The detailed collection of such identified risks allows them to be analyzed and reduced later. This helps the governance work to be monitored in real-time and assists in continuous compliance assurance work, also feeding system fine-tuning and improvements.

6. CONCLUSIONS

The co-design aspect of embedding trustworthy AI by using a process such as Z-Inspection® offers a dynamic and adaptable framework for addressing the ethical challenges posed by rapidly evolving AI technologies. Unlike static checklists or rigid compliance measures, this approach allows for incorporating new ethical considerations, technological advancements, and societal shifts as they emerge.

The methodology's flexibility is particularly valuable in addressing novel ethical challenges posed by cutting-edge AI technologies and responding to changing regulatory landscapes and societal expectations. By embedding diverse perspectives and continuous ethical reflection into the fabric of AI development, Z-Inspection® represents a shift in AI ethics and governance.

This approach goes beyond traditional assessment frameworks by integrating co-design principles throughout the AI system lifecycle. It facilitates collaboration among diverse stakeholders in the organization that are otherwise often working in silos. As co-design starts at the inception of AI system design it can facilitate a reduction of tensions between the areas of expertise in the organization. This interdisciplinary cooperation improves the comprehensive understanding of potential impacts and risks, leading to an improved buy-in of AI technology within the organization.

The iterative nature of the Z-Inspection® process promotes continuous evaluation and refinement throughout the AI lifecycle, treating ethical, societal and legal considerations as integral components rather than afterthoughts. This proactive approach enables early risk identification, real-time feedback on system design, and ongoing alignment with social responsibility initiatives and market expectations.

By fostering transparency and open dialogue, the co-design approach builds trust among all parties involved, including potential end-users and the broader public. It also allows for the identification and monitoring of risks that may not be immediately resolvable, supporting ongoing governance and compliance efforts.

In essence, the co-design aspect of Z-Inspection® offers a promising path toward creating human-centric AI systems that are not only technically robust but also ethically sound, socially beneficial, and regulatory compliant. This holistic approach to AI development and assessment is crucial in ensuring that AI technologies align with human values and societal needs as they continue to advance and integrate into various aspects of our lives. This approach also strives to promote the use of innovative technology and to improve design process maturity.

[13] Düdder, B., F. Möslein, N. Stürtz, M. Westerlund, and R. V. Zicari, 2021, "Ethical maintenance of artificial intelligence systems," in Pagani, M., and R. Champion (eds.), Artificial Intelligence for Sustainable Value Creation. Edward Elgar Publishing

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