DIGITAALINEN YHTEISKUNTA, IN ENGLISH 24.4.2024

Towards the co-design, assessment, and validation of trustworthy AI within the MANOLO project

TEKSTI | Paulinus Ofem , Emilia Risu , Sari Sarlio-Siintola

Artificial intelligence (AI) ignites curiosity and concern in equal measure, driving conversations about its vast potential and inherent risks. For Laurea University of Applied Sciences, the Ethics Guidelines for Trustworthy AI (AI HLEG 2019) have proven invaluable in navigating the complex ethical landscape of AI development in numerous research and development projects. The MANOLO project, with its Z-inspection® process (Zicari et al. 2021), represents a natural progression toward a more systematic and collaboratively driven ethical & legal analysis of AI-based solutions and their use.

Figure 1: Group photo for the MANOLO project kick-off in Dublin in February 2024 (Source: Q-Plan, MANOLO Consortium Member)

Towards AI Governance: Existing Initiatives and Frameworks

In recent years, AI capabilities have expanded rapidly, hence the quest for AI trustworthiness. These capabilities have been witnessed in critical sectors such as healthcare (Vetter et al. 2023; Rajamäki et al. 2023), manufacturing (Zeba et al. 2021; Kim et al. 2022), and telecommunications (Balmer et al. 2020). General stakeholders' concerns, ranging from data privacy and decision-making biases to job losses induced by the introduction of robots, caused governments, including private and public organisations and researchers, to proffer ways to govern AI technologies and their responsible use. To reap the benefits of AI and stem their perceived risks, Trustworthy AI presents itself as a vantage pathway, ensuring that AI technologies respect fundamental human rights while serving the public good transparently. In this vein, the European Commission established a High-Level Expert Group on Artificial Intelligence (AI HLEG 2019) to chart the course of Trustworthy AI. The AI HLEG was mandated to define trustworthy AI and propose strategies for implementing trustworthy AI across the European Union.

The European Commission ethics guidelines for Trustworthy AI (AI HLEG 2019) established seven critical requirements for Trustworthy AI: 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 well-being, and 7) accountability. These key pillars of the guideline aim to encourage lawful, ethical, and robust AI that is good for people and society. Apart from AI HLEG's guidelines, the Organisation for Economic Co-operation and Development (OECD)also established principles for Trustworthy A. Similarly, the OECD opines that AI systems should be developed in ways that respect the rule of law, human rights, democratic values and diversity (OECD 2019). OECD guidelines emphasise growth, sustainable development and society's well-being.

Therefore, it is common knowledge that AI technologies must be evaluated for their trustworthiness before their deployment in production environments. The Z-Inspection® Initiative is one of several initiatives that has leveraged AI HLEG's guidelines to create a framework (Zicari et al. 2019) for evaluating AI trustworthiness. It consists of three phases: a) Set-up phase (validation of pre-condition before commencement of actual Trustworthy assessment), b) Assess phase (creation and analysis of socio-technical scenarios and identification of ethical issues and validation of claims), and c) Resolving phase (addresses ethical tensions and makes recommendations).

The Z-Inspection® framework applies to all the phases of AI development, including design, development, deployment, monitoring and decommissioning. It has proved helpful in evaluating AI trustworthiness in several use cases, including health, as provided in (Zicari et al. 2019; Vetter et al. 2023).

Towards Trustworthy Efficient AI for Cloud-Edge

Computing: The MANOLO Project

The Laurea University of Applied Sciences is part of the MANOLO project, which aims to provide Trustworthy, Efficient AI for Cloud-Edge Computing. The project involves 18 partners from eight countries, and the European Union funds it under Grant Number 101135782.

"The overall vision of MANOLO is to deliver a complete and trustworthy stack of algorithms and tools to help AI systems achieve better efficiency and seamless optimisation in their operations, resources and data required to train, deploy and run high-quality and lighter AI models in both centralised and cloud-edge distributed environments" (MANOLO GA 2023, 2). Fig. 2 summarises MANOLO's key objectives.

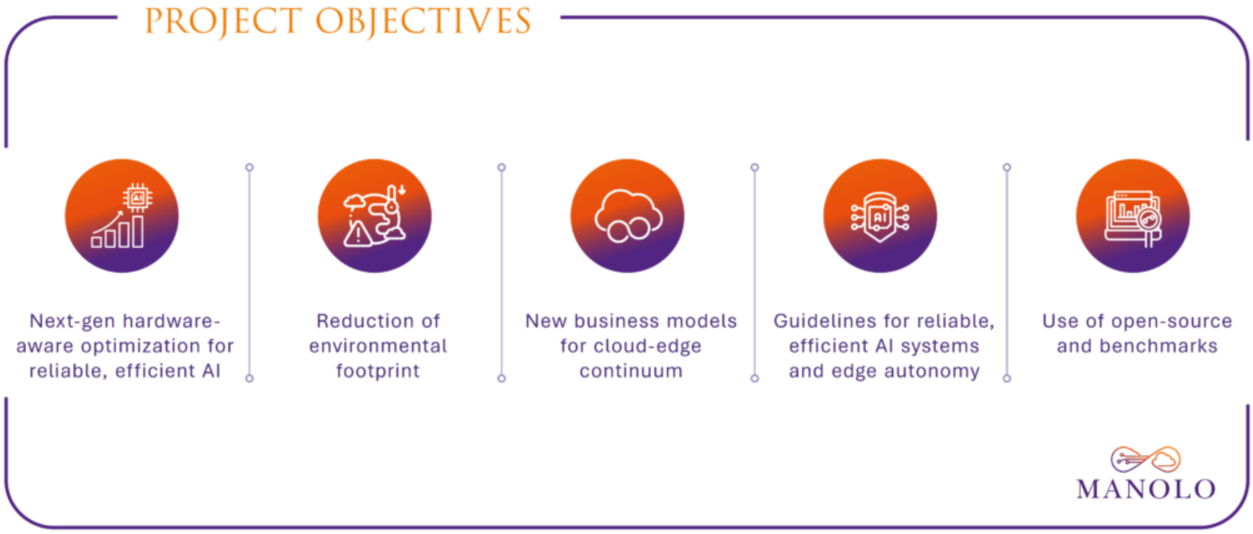

PROJECT OBJECTIVES
Next-gen hardware-aware optimization for reliable, efficient AI
Reduction of environmental footprint
New business models for cloud-edge continuum
Guidelines for reliable, efficient AI systems and edge autonomy
Use of open-source and benchmarks
MANOLO

Figure 2: MANOLO project objectives (Source: Q-Plan, MANOLO Consortium Member)

Fig. 3 presents the MANOLO overall concept as promised in the grant agreement. The concept is predicated on developing algorithms for training, understanding, compressing and optimising machine learning models (MANOLO GA 2023). The HW-aware Model Training and Optimisation component is at the heart of the concept, where a stack of algorithms is developed and deployed for different functions. Another critical component of the MANOLO concept is the Cloud Edge Continuum AI Model Function Allocation, which refers to strategies (e.g., the Cloud-edge Resource and Infrastructure Mapping module that collects, organises, and maps information) used to allocate resources to various AI learning and training needs. As the name implies, the Data Inspection and Generation component considers data collection, models and context from MANOLO use cases. This component will enable the development, testing, and validation of MANOLO algorithms for trustworthiness. The complete assessment of trustworthiness and monitoring of MANOLO outcomes is embodied in the Trustworthy Efficiency and Performance Benchmarking component. Lastly, the Human Oversight Trustworthy Assessment component deals with the process and tools for conducting the assessment and monitoring.

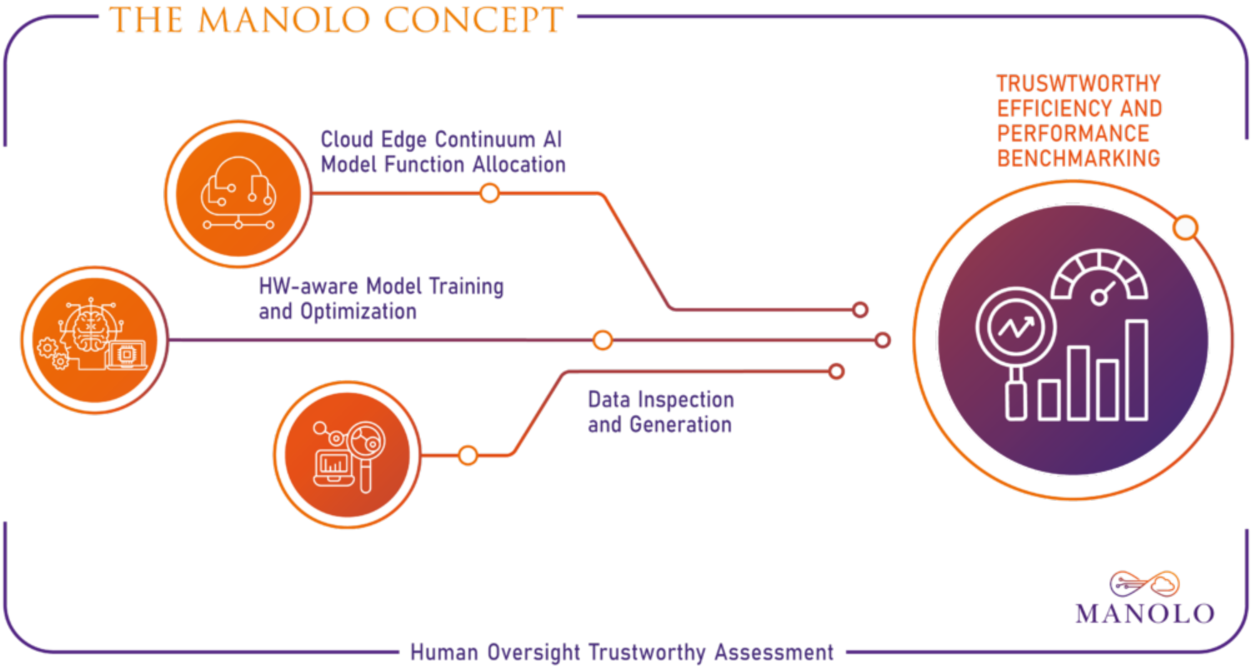

THE MANOLO CONCEPT
Cloud Edge Continuum AI Model Function Allocation
HW-aware Model Training and Optimization
Data Inspection and Generation
TRUSWTWORTHY EFFICIENCY AND PERFORMANCE BENCHMARKING
MANOLO
Human Oversight Trustworthy Assessment

Figure 3: The MANOLO overall concept. (Source: Q-Plan, MANOLO Consortium Member)

AI trustworthiness evaluation mechanisms will be embedded in the MANOLO solutions, using the Z-Inspection® methodology for Trustworthy AI assessment. This approach will also help AI systems conform to AI regulations. The MANOLO trustworthy AI evaluation process will consider the recently approved Artificial Intelligence Act (Artificial Intelligence Act 2024), which establishes a framework for the ethical use and provision of AI systems in the EU.

The MANOLO outcomes will be deployed as a toolset and tested in lab environments with use cases (see Fig. 4) using different distributed AI paradigms within cloud-edge continuum settings. Use cases will be validated in domains such as health, manufacturing, and telecommunications, aligned with market opportunities, as well as devices like robotics, smartphones, IoT, and neuromorphic chips (MANOLO GA 2023).

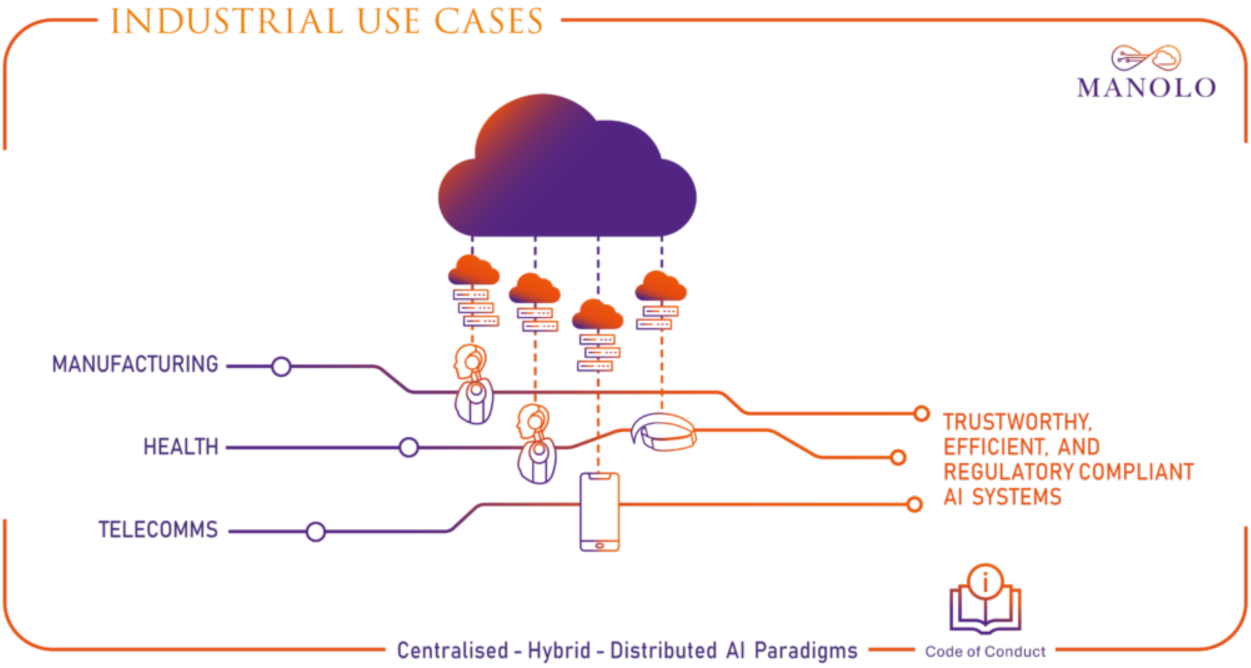

INDUSTRIAL USE CASES
MANOLO
MANUFACTURING
HEALTH
TELECOMMS
TRUSTWORTHY, EFFICIENT, AND REGULATORY COMPLIANT AI SYSTEMS
Centralised - Hybrid - Distributed AI Paradigms
Code of Conduct

Figure 4: MANOLO project use cases. (Source: Q-Plan, MANOLO Consortium Member)

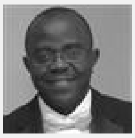
Funded by the European Union under GA no. 101135782. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or CNECT. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

References

- Artificial Intelligence Act: MEPs adopt landmark law. 2024. Accessed 22.3.2024.
- Balmer, R.E., Levin, S.L. and Schmidt, S. 2020. Artificial Intelligence Applications in Telecommunications and other network industries. Telecommunications Policy, 44(6), p.101977.
- High-Level Expert Group on Artificial Intelligence (AI HLEG). 2019. Ethics guidelines for trustworthy AI European Commission. Accessed 18.3.2024.
- High-Level Expert Group on Artificial Intelligence (AI HLEG). 2020. Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. European Commission. Accessed 23.3.2024.
- Kim, S.W., Kong, J.H., Lee, S.W. and Lee, S. 2022. Recent advances of artificial intelligence in manufacturing industrial sectors: A review. International Journal of Precision Engineering and Manufacturing, pp.1-19.
- MANOLO Grant Agreement (MANOLO GA). 2023.
- Organisation of Economic Co-operation and Development (OECD) Principles on Artificial Intelligence. 2019. Accessed 23.3.2024
- Rajamäki, J., Gioulekas, F., Rocha, P.A.L., Garcia, X.D.T., Ofem, P. and Tyni, J. 2023, May. ALTAI Tool for Assessing AI-Based Technologies: Lessons Learned and Recommendations from SHAPES Pilots. In Healthcare (Vol. 11, No. 10, p. 1454). MDPI.
- Vetter, D., Amann, J., Bruneault, F., Coffee, M., Düdder, B., Gallucci, A., Gilbert, T.K., Hagendorff, T., van Halem, I., Hickman, E. and Hildt, E. 2023. Lessons learned from assessing trustworthy AI in practice. Digital Society, 2(3), p.35.
- Zeba, G., Dabić, M., Čičak, M., Daim, T. and Yalcin, H. 2021. Technology mining: Artificial intelligence in manufacturing. Technological Forecasting and Social Change, 171, p.120971.
- Zicari, R.V., Brodersen, J., Brusseau, J., Düdder, B., Eichhorn, T., Ivanov, T., Kararigas, G., Kringen, P., McCullough, M., Möslein, F. and Mushtaq, N. 2021. Z-Inspection®: a process to assess trustworthy AI. IEEE Transactions on Technology and Society, 2(2), pp.83-97.

URN http://urn.fi/URN:NBN:fi-fe2024042421447



Paulinus Ofem

paulinus.ofem(at)laurea.fi

Project Specialist

Emilia Risu

emilia.risu(at)laurea.fi

Service Designer

Sari Sarlio-Siintola

sari.sarlio-siintola(at)laurea.fi

Senior Lecturer

AI governance

Ethics governance

MANOLO project

Trustworthy AI

Z-Inspection®