Raising Awareness on Ethical Principles Applied to OPTIMAI

UAB

Universitat Autònoma de Barcelona
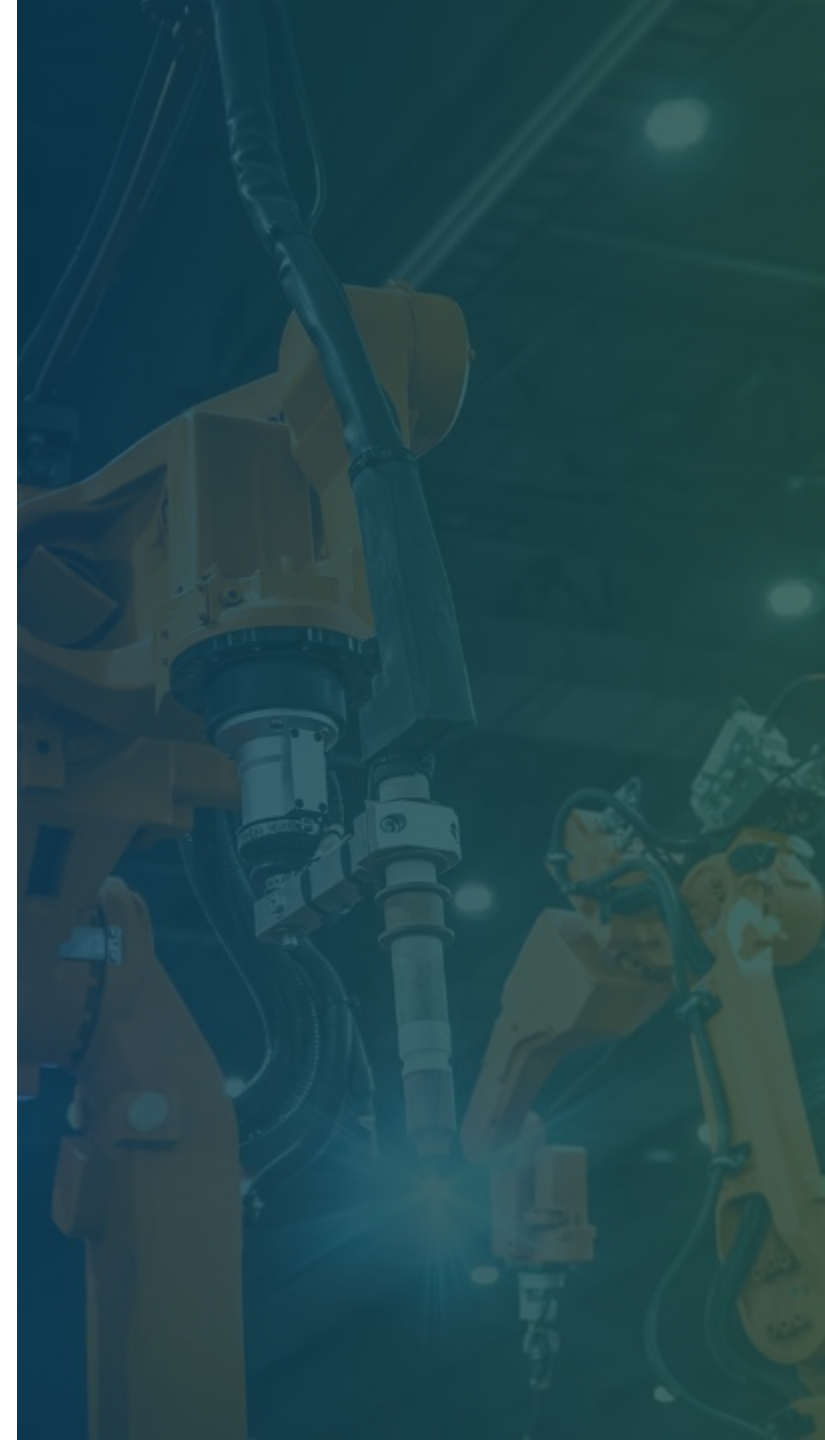
OPTIMAI

Contents

OPTIMAI

Introduction

1

Introduction

OPTIMAI piloting activities are subject to be conducted in strict compliance with the ethics research framework identified as applicable for the OPTIMAI Project.
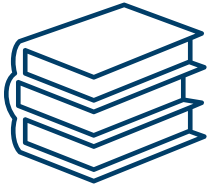
It comprises the highest standards, principles, and good practices of research ethics stated in:

i)   the European Code of Conduct for Research Integrity [1];

ii) the EU Ethical Responsible Research and Innovation Framework (RRI) [2];

iii) Responsible Research Innovation in Industry [3]; and

iv) Ethical principles and guidelines for trustworthy Artificial Intelligence (AI) [4].

Ethical AI Trustworthy Principles

This ethical training focuses on the ethical AI trustworthy principles identified as applicable to OPTIMAI.

These AI ethical principles are:

1   **Human Autonomy**

2   **Prevention of Harms**

3   **Fairness**

4   **Explicability**

and will be operational in OPTIMAI piloting activities as follows:

⊘PTIMAI

Human Autonomy

2

Human Autonomy

Human autonomy is a central consideration when creating ethical AI-enabled technologies.

The principle of human autonomy implies that AI-enabled technologies should be designed and deployed in a way that respects and protects fundamental rights and ensures human agency and oversight.

This means that AI-enabled technologies should not be designed to subordinate, coerce, deceive, manipulate, condition or herd humans, but to augment, complement and empower humans.

OPTIMAI
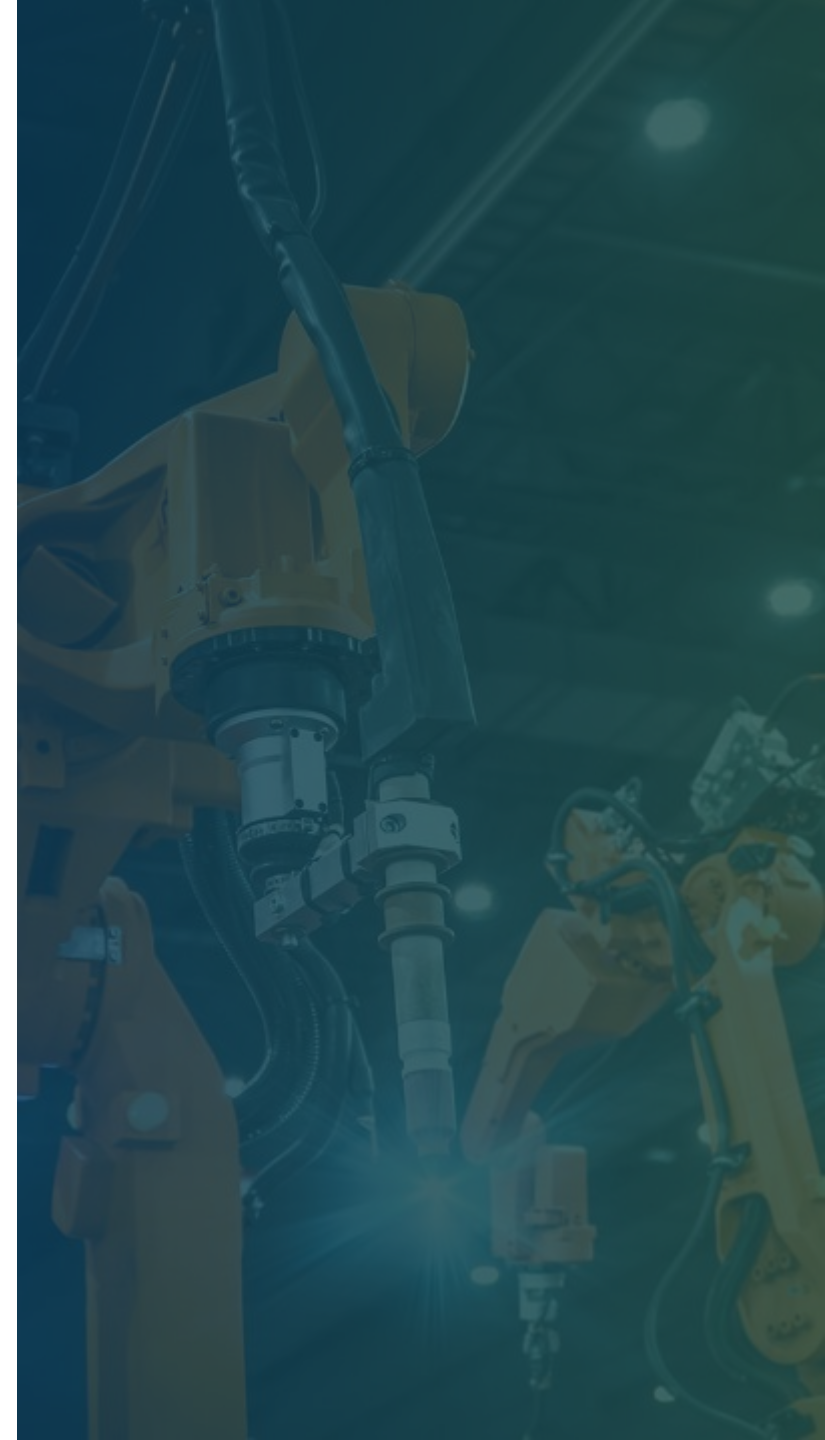
Human Dignity

AI-enabled technologies must ensure human dignity.

In the workplace, the objectification and dehumanisation of workers must be avoided. In this regard, the physical and mental health of workers must be protected.

Worker's dignity might also be undermined by the consequences that the deployment of AI systems in the workplace may have on the de-skilling of the labour force and the meaning of work.

In OPTIMAI, the technological solutions are being designed and deployed to help workers in their daily tasks and to up-skill them; in no way is OPTIMAI conceived to replace workers with machines.

OPTIMAI

Voluntariness

To protect worker's dignity, ensuring voluntariness is key.

Given the power imbalance in the workplace, workers may feel coerced to use AI systems in the workplace or may fear consequences if they refuse to adopt them.

In OPTIMAI, the voluntary participation of workers in the piloting activities will be based on an Informed Consent Form and an Information Sheet, in which workers will be informed about the OPTIMAI project, the piloting activities and their rights.

Before their participation in the piloting activities, workers must read the Informed Consent Form and the Information Sheet, and they must be given the time and opportunity to ask questions. The Informed Consent Form must be signed by the worker and by the pilot site responsible.

OPTIMAI

Informed Decision-Making

To ensure human agency, workers should be able to make informed autonomous decisions regarding the outcomes of the OPTIMAI solutions and should have the skills to assess and challenge such outcomes.
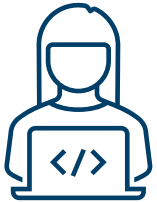
Therefore, training sessions must be delivered to ensure that workers have the knowledge to understand how the OPTIMAI solutions work and how to interact with them.

Risk Minimisation

The purpose of human oversight is to prevent or minimise the risks that technological solutions may pose.

In this regard, meaningful human control can only be achieved if human-centric design principles and appropriate human-machine interfaces are embedded into the technologies.

Additional measures should be implemented to ensure that workers have the expertise, necessary competencies, and authority to exercise human control effectively. For instance, this can be achieved through training sessions that enable the understanding of the capacity and limitations of the OPTIMAI solutions.

OPTIMAI

Prevention of Harms

3

Prevention of Harms

The principle of prevention of harms means that technological solutions should neither cause harm to individuals or groups nor exacerbate it or have negative impacts on human beings.

In the workplace, this implies that workers' dignity must be respected, and their mental and physical integrity protected.

To minimise the impact of OPTIMAI solutions on workers, a participatory approach is encouraged to be adopted where workers are involved in the design, development, and deployment of the technology so that it can be adapted to their needs.

OPTIMAI
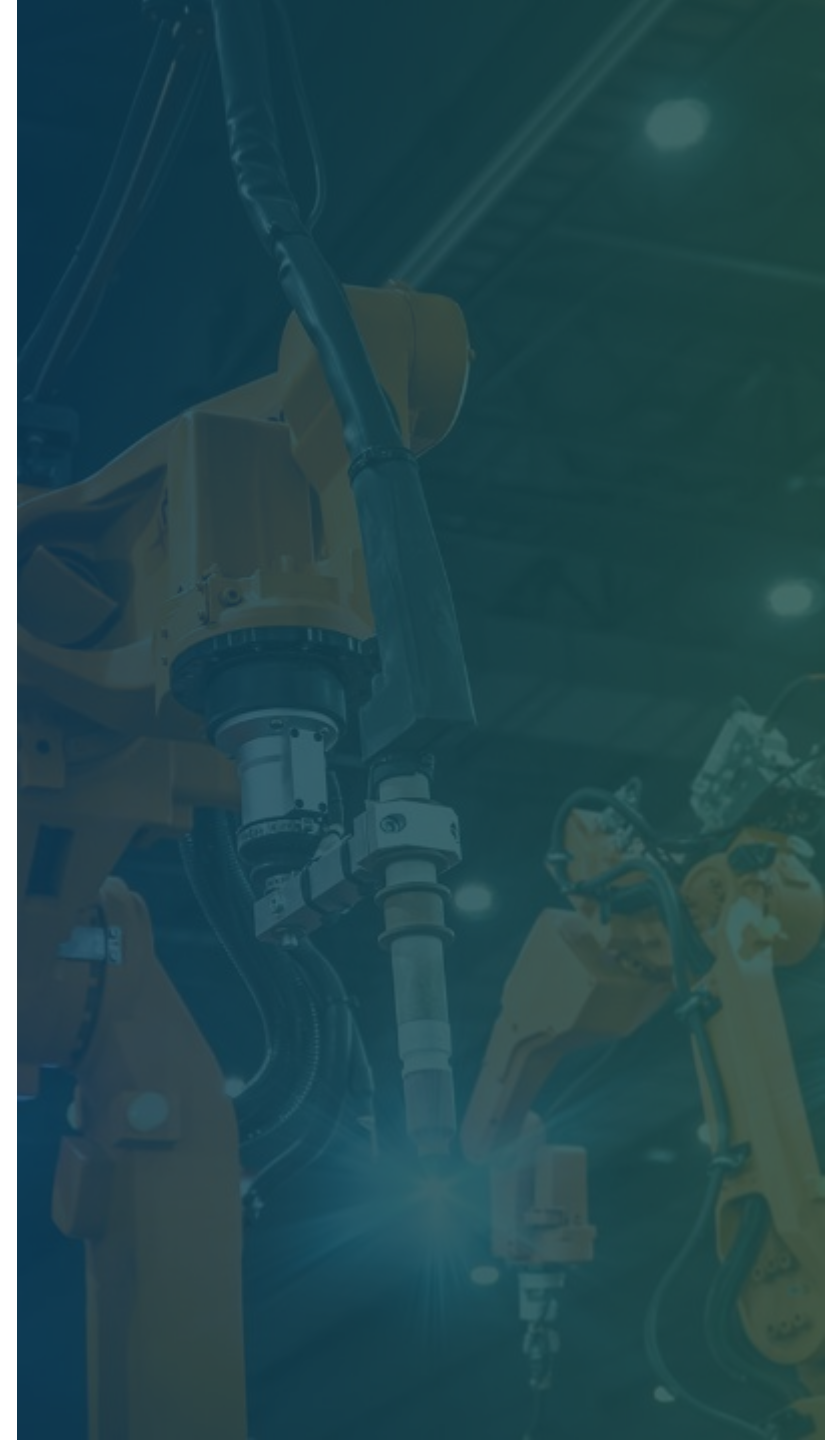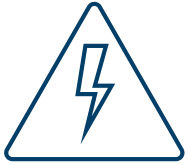
Potential Harms

The potential harms that can be caused by technological solutions also require addressing:

i)   the technical robustness and safety of the technology;

ii) privacy and data governance concerns; and

iii) societal and environmental well-being.

OPTIMAI

Technical Robustness & Safety of Technology

Firstly, this means that OPTIMAI solutions must be robust, resilient, secure, safe, accurate, reliable, and reproducible.

Technical robustness and resilience should be ensured to prevent the exploitation of vulnerabilities by third parties and misuse.
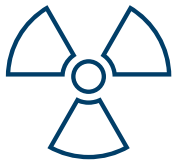
Therefore, the existence of potential security risks must be evaluated at the design, development and deployment phases, and mitigation measures must be implemented in accordance with the magnitude and likelihood of the risks.

OPTIMAI

Technical Robustness & Safety of Technology

Security and safety measures should also be put in place to enhance workers safety and prevent risks.

OPTIMAI solutions must also be accurate. Accuracy rates should be particularly high when such systems can directly affect individuals. Accuracy must be monitored on an ongoing basis and procedures to mitigate and correct potential risks must be implemented.

Additionally, workers need to trust the system to use it, therefore reliability and reproducibility are key aspects to ensure the adoption of the technology among workers.

OPTIMAI

Privacy & Data Governance Concerns

Secondly, the prevention of harms to privacy and data protection is paramount given the potential risks that AI-enabled technologies pose to these fundamental rights through the processing of massive amounts of personal data.

Importantly, in OPTIMAI, the processing of personal data is not intended. However, it is worth mentioning that OPTIMAI partners must respect workers' right to privacy and data protection by complying with the GDPR and by aligning with existing standards or widely adopted protocols.

In IoT environments, it is particularly crucial to clarify data ownership, the roles of data controllers and processors and access to data.

Oversight mechanisms must also be put in place to ensure data quality and integrity that minimises the risks of using biased, inaccurate, or compromised datasets. Therefore, processes and datasets must be scrutinised and documented.

OPTIMAI

Societal and Environmental Wellbeing

Lastly, the use of AI-enabled technologies should aim at benefitting society and the environment.

AI systems must be designed, developed, and deployed with sustainability and environmental friendliness in mind. Therefore, the ecological impact of the OPTIMAI solutions should be evaluated and measures to reduce such impact should be implemented.

Societal and Environmental Wellbeing

The social impact of the system should be regularly assessed both at the individual and societal level to ensure that OPTIMAI solutions do benefit workers and the society as a whole.

For instance, the evaluation of the impact of the technology on workers should cover physical and mental health issues, non-discrimination, de-skilling of the workforce, among others.

As for the societal considerations, the impact on the job market and the societal consequences it may entail should be evaluated and, if necessary, mitigation measures should be provided.

OPTIMAI

Fairness

4

Fairness

The principle of fairness entails equality, diversity and the prevention of discrimination and stigmatisation against individuals and groups.

Fairness can be achieved by:

i)   promoting diversity, inclusion and non-discrimination;

ii) fostering societal and environmental well-being while reducing potential harms; and,

iii) adopting accountability measures.

OPTIMAI

Promoting Diversity, Inclusion & Non-Discrimination

Firstly, diversity and non-discrimination can be enhanced with oversight processes that identify, examine, address and test biases in the datasets and at the design and development phases of the OPTIMAI solutions.

From a design perspective, technology should be understandable and accessible to all workers regardless of their age, abilities, or characteristics.

In this regard, the participation of relevant stakeholders with diverse backgrounds and viewpoints at the different stages is highly encouraged.

In the workplace, impacted workers and their representatives can be engaged in such discussions.

OPTIMAI

Fostering Societal and Environmental Wellbeing & Harm Reduction

Secondly, OPTIMAI solutions should be designed to strive for social and environmental well-being.

Concerning the principle of fairness, the social impact of the OPTIMAI solutions on workers should be evaluated in terms of causing or exacerbating discrimination, stigmatisation, or marginalisation, and if they are having negative effects on the skilling of the workforce

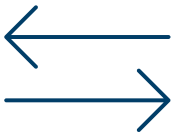OPTIMAI

Adopting Accountability Measures

Lastly, accountability requires the implementation of appropriate technical and organisational measures to report the system's performance and provide effective remedy and redress to the extent possible.

Auditability involves reporting the negative impacts of the system, identifying appropriate mitigation measures, and feeding them into the system.

Accountability also includes providing explanations of the system's outcomes and the ability to seek redress.

To this end, internal communication channels can be established for workers to submit their complaints, without risk of retaliation, and seek redress for harms caused by the OPTIMAI solutions.

OPTIMAI

Explicability

5

Explicability & Traceability

The principle of explicability requires transparency of the OPTIMAI solutions – including the datasets and their inner workings – which ultimately enables human oversight.

For OPTIMAI solutions to be transparent, traceability measures must be implemented. This implies that datasets and the technology that underlies the technological solution should be documented.

Given that traceability allows for the identification of the reasons behind the solutions' outcomes, it enables explainability.

OPTIMAI

Explainability

Explainability means the ability to explain the outcomes made by the technological solution intelligibly.

In this regard, communication is crucial since workers must be aware that they are interacting with an AI system in the first place in order to be able to request an explanation.

Consequently, workers must be informed in a clear and plain manner about their interaction with the OPTIMAI solutions, how they work and their purpose, as well as their capabilities and limitations.

References

6

References

[1]   https://allea.org/code-of-conduct/

[2]   https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation;

      https://rri-tools.eu/en/about-rri

[3]   https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

      https://www.unglobalcompact.org/

      https://mneguidelines.oecd.org/mneguidelines/

      https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_ent/---multi/documents/publication/wcms_094386.pdf

      https://www.ilo.org/declaration/thedeclaration/textdeclaration/lang--en/index.htm

      https://www.iso.org/iso-26000-social-responsibility.html

[4]   https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

OPTIMAI

OPTIMAI

Thank you

UAB
Universitat Autònoma
de Barcelona