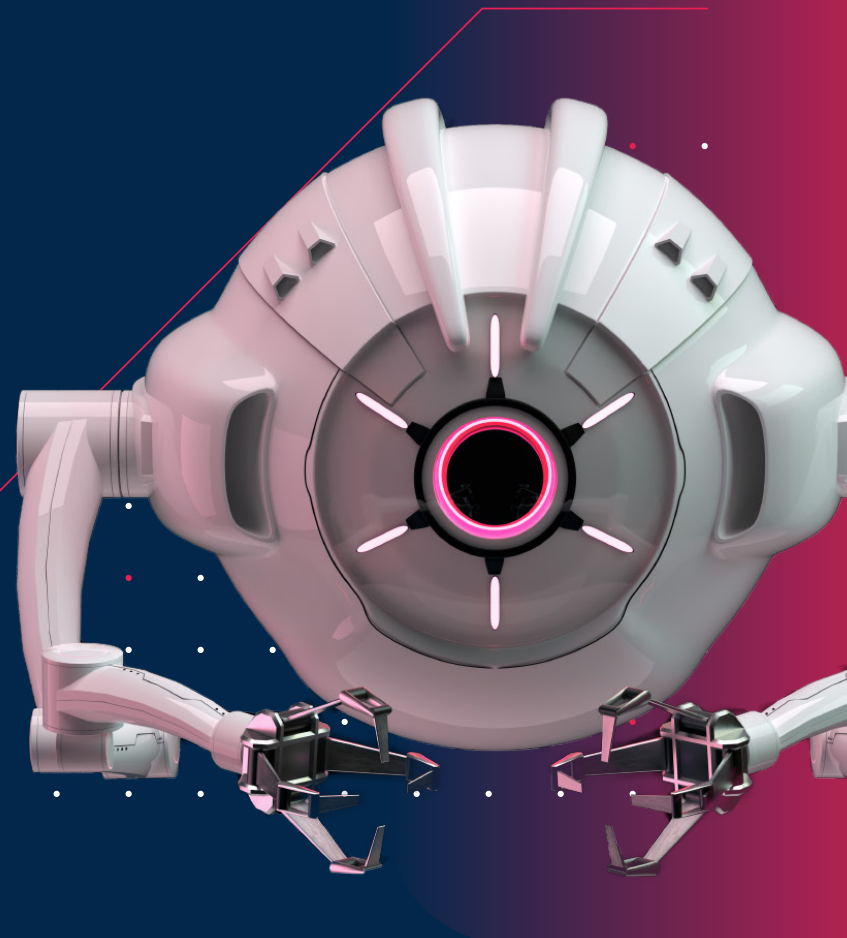
Robotics4EU
Boosting Wider Adoption of Robotics in Europe

Factsheet #7.

The Ethical Dimensions of Responsible Robotics

This document provides an overview of the ethical dimension of responsible robotics because while the rapid technological advancements in robotics and AI technologies promise countless benefits, they also raise profound ethical questions. Robots continue to evolve in complexity, being able to make autonomous decisions that influence people and their well-being.

Disclaimer
This factsheet is based on research conducted by the Robotics4EU, as well as second-hand data.



What are the ethical and moral dimensions relevant to creating and using robots?

Privacy and Surveillance

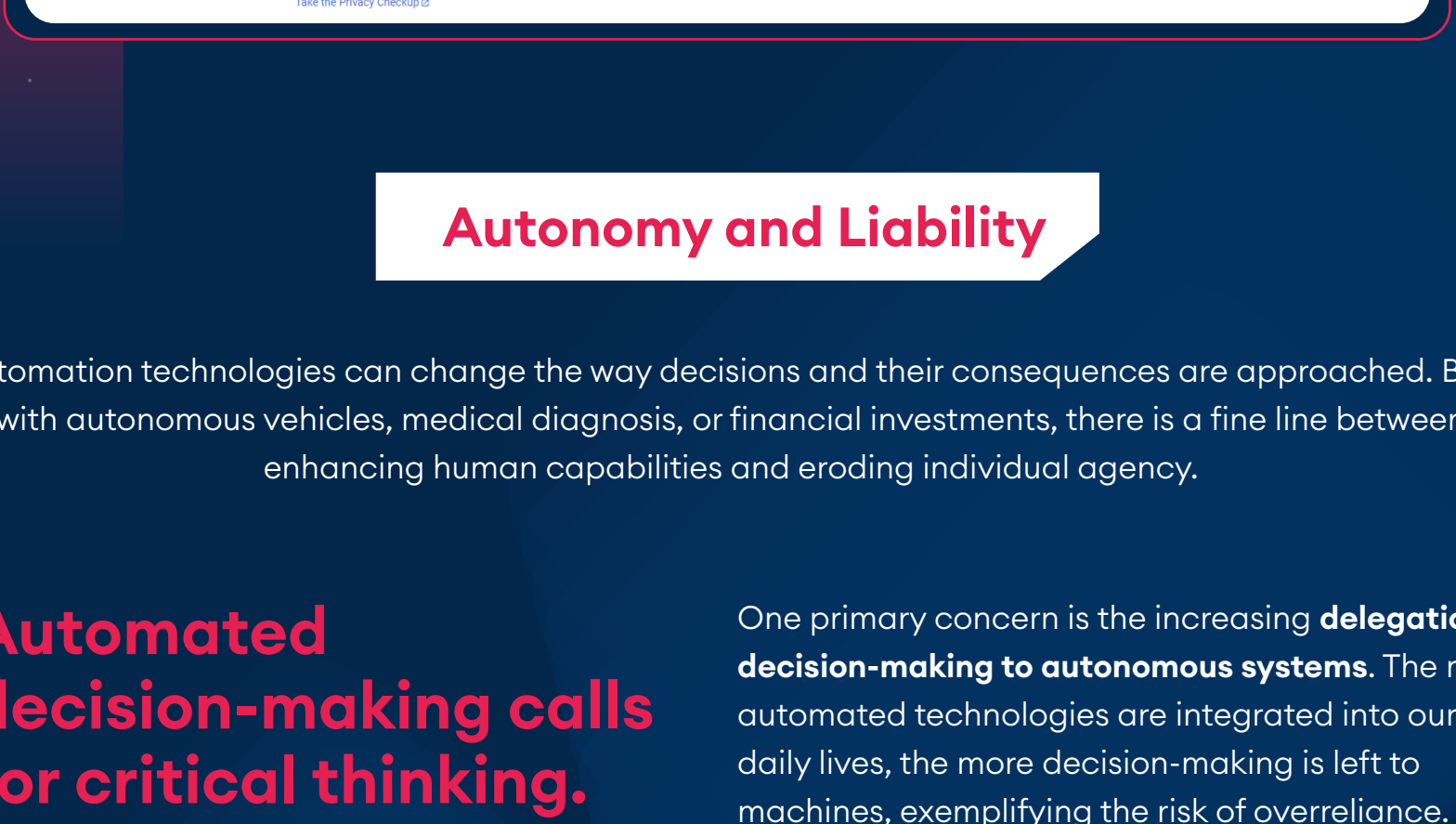
AI systems and robots are increasingly equipped with sophisticated sensors and cameras with advanced data-processing capabilities. Due to the integration into the growing Internet of Things, where information is shared beyond one device or location, the boundary between public and private spaces becomes blurred. **These technologies offer convenience and security benefits, but they also raise public privacy concerns.**

Robotics and AI-driven **surveillance systems** can capture a wealth of information about individuals, from their daily routines and behaviours to their personal preferences and intimate moments. Harnessed without proper safeguards, **data can be exploited** for targeted advertising, social manipulation, government surveillance, or other purposes. Moreover, the vulnerability of these systems to data breaches and cyberattacks - if not properly addressed - poses substantial risks to individuals' privacy and data security. To tackle this challenge and to find the right equilibrium, solid **legal frameworks**, universal data protection measures, and **open public discussions** are required.

The 'constantly active' nature of automated systems also raises questions about consent. In some cases, individuals may not be aware of the extent to which data is collected, let alone have the **opportunity to give or withdraw consent**. Furthermore, privacy policies are often indecipherable. To improve the acceptability of new technologies, their ways of operating (especially with data) must be **made available and clear to the users** at any point in time. The development and deployment of robotic systems must therefore prioritize the protection of personal freedoms, keeping communication on this topic as transparent as possible.

Twitter/X and Google have good examples of transparent, well-described and user-friendly privacy policies (bonus points for embedded tutorial videos on how to change data and privacy settings).

See also the Robotics4EU report on data and privacy.



Autonomy and Liability

Automation technologies can change the way decisions and their consequences are approached. Be it with autonomous vehicles, medical diagnosis, or financial investments, there is a fine line between enhancing human capabilities and eroding individual agency.

Automated decision-making calls for critical thinking.

One primary concern is the increasing **delegation of decision-making to autonomous systems**. The more automated technologies are integrated into our daily lives, the more decision-making is left to machines, exemplifying the risk of overreliance.

To avoid losing human autonomy and compromising the ability to handle unforeseen situations, **critical thinking** and other skills **necessary to cope without the use of technologies** must be taught at all levels of education.

Regulations help to mitigate possible harm caused by robots.

A major question has forever inspired sci-fi authors, philosophers, law scholars and many more: who should be held responsible if robot action results in harm – the manufacturer, the programmer, the operator, or the autonomous system itself? Determining liability/responsibility is essential not only for legal purposes but also for ensuring that appropriate corrective actions are taken to **prevent future harm**. This calls for policymakers to develop and enforce **clear rules for responsibility and accountability**, which consider different contexts, robots' level of autonomy and the potential consequences of their decisions.

Research and promote better explainable AI

In 2024, **understanding the decision-making process of autonomous systems is still complicated**, especially in machine learning – the capacity of robots to make decisions far exceeds their capacity to explain these decisions. Developing and **promoting transparent mechanisms to record and explain autonomously made decisions** is therefore needed. One way to encourage to encourage the application of research into explaining AI is to organize challenges with a scoring system that prioritizes the explainability of AI, such as the RoboCup.

Some resources for developers of AI algorithms:

Open source explainable AI toolkit:
https://xaitk.org/

European ethics guidelines for trustworthy AI:
https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

TTC joint roadmap for trustworthy AI:
https://digital-strategy.ec.europa.eu/en/library/ttc-joint-roadmap-trustworthy-ai-and-risk-management

NIST resources and research on Explainable AI:
https://www.nist.gov/artificial-intelligence/ai-fundamental-research-explainability

See also the Robotics4EU report on legal dimension.

Cybersecurity

The proliferation of robots in various domains has **increased connectivity and reliance on software and data**, making robots more attractive targets for cyberattacks. The main cybersecurity concerns are **unauthorized access** and control, data **theft**, and privacy breaches. Attackers may exploit vulnerabilities to **manipulate robots**, steal sensitive information, or disrupt their operations, but also compromise robot security through insider threats and supply chains. This includes **Denial of Service (DoS)** attacks that make systems crash and critical applications in healthcare or autonomous vehicles unavailable.

To mitigate cybersecurity risks, **security measures** should include secure software and hardware development, access controls, data encryption, and regular updates, but also following **industry standards** such as ISO 31000 throughout the development and use of the robot. In many cases, operators' awareness and know-how of the right (or best) practices are vital, so **user education and training are as important** as technological risk mitigation.

The challenges and consequences of cybersecurity **threats vary depending on the area of application – the measures of protection therefore need to be adapted** according to the concerns of a specific area.

For **industrial robots**, the consequences could include destruction of property, loss of revenue, or even industrial espionage, which could be prevented with strategies including compartmentalization of data and dataflow testing.

For a **healthcare system**, a failure in cybersecurity could cause loss of an adapted treatment, direct harm to a patient, or sensitive data leaks while the proper mitigation steps would need to include isolation of critical infrastructures from redundant systems such as paper backups of patient data.

Bias and Discrimination

Finally, a major defect of AI systems is one that robots share with people: **the reproduction of biases due to biased input data**. If the initial dataset is not corrected for existing societal biases, **AI can unintentionally perpetuate and exacerbate** inequalities in its results.

If training data contains inherent biases, **such as gender-based pay disparities or racial profiling**, these biases are encoded into the algorithms. Consequently, robots would make decisions that can discriminate against certain genders or racial groups – such biased AI systems can lead to unfair hiring practices, where qualified candidates are rejected due to **algorithmic prejudice.**

As with previous aspects, the opacity of AI algorithms makes it challenging to detect and rectify bias. Without **transparency and accountability mechanisms**, it is difficult to ensure that robots make fair and unbiased decisions. Setting the focus on **explainable robots** and enforcing **quality guidelines** for training and testing datasets will help to move forward with more human-friendly robots.

Good practices for inclusivity and mitigation of discrimination include:

Testing datasets that are specifically unbiased and include a fair proportion of minority representatives. This not only improves the performance of minority groups, but if done properly can improve the performance of the system over all groups.

User-centered design including a variety of end-users in the design process, particularly for physical systems such as robots that can create barriers to access for certain groups:
https://www.frontiersin.org/articles/10.3389/frobt.2022.731006/full

Datasets and methodology designed to measure and mitigate **Large Language Model bias** across different characteristics:
https://ai.meta.com/blog/measure-fairness-and-mitigate-ai-bias/

consortium

CIVITTA
robotex International
LOBA
LNE
Fonden Teknologirådet
AgriFood DIH Lithuania
NTNU Norwegian University of Science and Technology

follow us on
@robotics4eu

contact us
info@robotics4eu.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017283