European Sociological Association

ESA RN24 ESA RN36

International Conference

Technology, Artificial Intelligence and Education Towards the Future of Learning

26 and 27 February 2026 - Link Campus University

Organised by RN24/SSTNET – Sociology of Science and Technology Network & RN36 – Sociology of Transformations: East and West.

Towards inclusive AI: regulatory and ethical challenges

Laura Evangelista
Concetta Fonzo
Marianna Forleo

INAPP
National Reference Point EQAVET

UNIVERSITÀ DEGLI STUDI LINK

AIS Teorie Sociologiche e Trasformazioni Sociali

Con il patrocinio di AIS Sociologia dell'Educazione

DiTES

Quaderni di Comunità

BUTH AI Building Trust in Human Centric Artificial Intelligence

Introduction

In a few decades, Artificial Intelligence (AI) has shifted from the realm of science fiction to a significant force influencing many fields of contemporary life with a strong impact on social, economic, and institutional relations.

DEFINITION of Artificial Intelligence

An accepted definition of artificial intelligence is lacking, but scholars refer to AI as the development of computer systems designed to perform **tasks that typically require human intelligence** (West, 2018).

Theoretical background

Symbolic approach	1950s and 1960s. It is based on the assumption that intelligence can be replicated through the manipulation of symbols and explicit logical rules (top-down approach). It failed to understand complex domains such as natural language (Alain Turing.)
Connectionist approach	Inspired by the biological structure of the brain (neural networks) and based on learning from data- rather than preprogramed rules (bottom-up approach).
Human centric approach	In regulation promoted by the European Union. AI should be a tool at the service of humanity, governed by ethical and legal principles (transparency, accountability) to protect fundamental rights.
Risk-based approach	The legislative methodology of the AI Act, which classifies technologies not according to their technical nature, but according to the potential harm they may cause to society.

- **Alan Turing**: Pioneer of AI concepts and the Turing Test.
- **John McCarthy**: Coined the term "Artificial Intelligence" and contributed to symbolic reasoning.
- **Geoffrey Hinton**: Key figure in deep learning and neural networks; known as the "Godfather of Deep Learning“, **Nobel Prize 2024.**

Over the years, AI has made significant progress, by operating through algorithms and statistical learning techniques on extensive datasets and developed **“Large Language Models (LLMs)”** which generate text that mimics human writing and supports various tasks.

AI rapid advancement, however, poses significant challenges to fundamental rights. AI-driven systems increasingly threaten **privacy** and contribute to **misinformation and discrimination**.

- **Bias and Discrimination:** amplify biases already present in their data and propagate harmful stereotypes and hegemonic narratives.
- **Data Colonialism:** Large AI companies extract massive amounts of data from human life, monitoring consumer behavior for commercial targeting purposes.
- **Mass Surveillance and Disinformation**: These systems can fuel the spread of false content, undermining public trust and democratic processes.

Research questions and objectives

How does the integration of AI technologies affect job displacement and creation across various sectors?

What are the ethical concerns associated with the deployment of AI in decision-making processes?

What regulatory frameworks are most effective in ensuring the ethical use of AI while promoting technological innovation?

Methodology

Documentary and comparative analysis.
Critical-descriptive approach, aimed at interpreting the impact of legal regulations on technological and social reality.

Research design: It’s a qualitative research which includes desk analysis and case study.

Case Study.
Use of real-world examples (ELIZA, COMPAS, Clearview AI, Biden's Robocall) to illustrate the practical application of theories of bias, privacy violation, and disinformation.

Main findings

Artificial Intelligence, Technological transformation as a

PARADOX:

- the potential to significantly improve human lives,
- substantial risks related to privacy, misinformation, and discrimination

Human Rights

Although AI technologies should improve the protection of human rights and law, they also present significant risks, which can lead to discrimination, gender inequality, threats to democratic processes, and human rights violations (Council of Europe 2024).

Legal framework

Until recently, authorities lacked the powers, procedural frameworks, and resources necessary to ensure and monitor compliance with AI regulations.

National bodies were responsible for enforcing safety standards and fundamental rights.

This legal uncertainty and complexity discouraged companies from dveloping and using AI technologies, slowing progress in Europe and reducing the EU's global competitiveness in this area.

Artificial Intelligence and the right to information

The rise of new Generative AI (GAI) technologies can manipulate and create several types of digital content easily enabling multi-modal misinformation; AI-driven personalization algorithms are powerful but also amplifiers of misinformation on social media platforms.

the “economics of attention”
an economic model in which investments are indirectly financed through advertising revenues.

The more time users spend on a platform, the more profit it makes.

Fake robocalls

By interacting with the platform, users inform the algorithm about their preferences, so that **the contents recommended closely align with users' interests and opinions**; this generates greater profits but, at the same time, the user ends up seeing and reading only a single perspective, given by the algorithm.

A key example of the application of Generative AI was the **spread of fake robocalls from Joe Biden in New Hampshire**.
Using Generative AI, a company had published counterfeit recordings of the US President's voice, suggesting that voters should not vote.

Artificial intelligence and the right to nondiscrimination

Generative AI operates through algorithms essential for streamlining numerous tasks that impact daily lives, often on a scale that exceeds human capabilities. Indeed, algorithms follow computer instructions to convert input into output. Despite these “superpowers,” the potential for bias in algorithms is extremely high.

Machines and technologies just appear impartial and neutral, in reality they are not. In fact, the people who develop algorithms may embody certain biases, which are easily transferred to machines.
Therefore, the tendency of algorithms to generate discriminatory outcomes is not intrinsic to AI technology itself but is rather deeply rooted in the social factors that influence those who train such technologies.

Artificial Intelligence and the right to privacy

The enormous amount of data created and exchanged online allows companies, governments, and organizations to improve their decision-making processes, but often, this data includes sensitive information that individuals would prefer to keep private.

PRIVACY, the right to keep personal information safe and secure from unauthorized access.

Due to the growing volume of personal data collected and analyzed, AI poses a risk to the privacy of both individuals and organizations, mainly due to the complexity of algorithms.
Furthermore, AI technology is capable of making decisions based on imperceptible data patterns that often escape humans.

EU Artificial Intelligence Act (2024)

It provides a **clear legal framework** for the development of AI, ensuring compliance with ethical and legal standards while promoting innovation in a responsible manner. Through these legal safeguards, the EU seeks to create a digital environment in which AI-driven services respect fundamental rights and promote fairness.

It is the world's first binding horizontal regulation on AI.

Definition of AI as a broad and technologically neutral system based on software that uses machine learning, logical and knowledge-based approaches, and statistical methods.

The EU AI Act regulates high-risk systems with transparency and accountability but urge global coordination, reliable AI with human oversight and a **shift from anthropomorphic trust to verifiable reliability**, drawing on EU guidelines and philosophical critiques to align innovation with human rights.

Critical Discussion

1. **AI is not neutral**, biases are not intrinsic flaws in the machine, but reflections of social inequalities present in the training data (Compass case)
2. The current economic model of platforms, the Economy of Attention, is structurally incompatible with the right to accurate information because personalization algorithms create disinformation and deepfakes. Furthermore, **mass data collection (web scraping) represents a systematic violation of privacy** that previous regulatory frameworks were unable to effectively contain (Clearview AI case).
3. From “Trust” to “Reliability”: this distinction **shifts the focus from morality to the responsibility of suppliers** and the need for constant technical controls. Abandoning the anthropomorphization of AI to adopt a model of controlled technical reliability
4. **Rejecting the idea that the negative impact of AI is inevitable**. The shape the technology will take in society will depend on the ability of institutions to impose limits on technocratic and commercial powers.
5. **Human Oversight**: it is not technically possible to develop AI that works flawlessly in all conditions.

Human oversight is not just an ethical requirement, but a functional necessity.

Conclusions

Over the years, AI has made significant progress, achieving the appearance of intelligent behavior through algorithms and statistical learning techniques on vast data sets.
This has enabled the development of "Large Language Models (LLMs)" such as those that generate human-like text.
However, this rapid progress poses significant challenges to fundamental rights.
Bias and Discrimination: LLMs often amplify biases present in training data, propagating harmful stereotypes.
Data Colonialism: Large AI companies mine massive amounts of human data, monitoring consumer behavior for commercial targeting purposes.
Mass Surveillance and Disinformation: These systems can fuel the spread of false content, undermining public trust and democratic processes.

The challenge for trustworthy AI is just at the beginning.
Since it's not yet possible to develop systems
operating flawlessly in all conditions,
human supervision remains essential and indispensable.

References

European Commission (2020), *White paper on artificial intelligence: A European approach to excellence and trust* COM(2020)

European Union (2024), *Artificial intelligence Act*, Official Journal of the European Union, L 1689, 12 July, https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Evangelista L., Fonzo C., Zecca E. (2025), An examination of fundamental rights protection in AI governance, Sinappsi, XV, n.2, pp.75-89

Mantelero A. (2022), *Beyond data: Human rights, ethical and social impact assessment in AI*, London, Springer Nature

McCarthy J. (2007), *What is Artificial intelligence*, Stanford (CA), Stanford University

Seow J.W., Lim M.K., Phan R. C.W., Liu J.K. (2022), A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities, *Neurocomputing*, 53, pp.351-371

Turillazzi A., Taddeo M., Floridi L., Casolari F. (2023), The digital services act: an analysis of its ethical, legal, and social
implications, *Law, Innovation and Technology*, 15, n.1, pp. 83-106

West D.M. (2018), What is artificial intelligence?, *Brookings.edu*, October 4

Thank you for your attention

Laura Evangelista l.evangelista@inapp.gov.it
Concetta Fonzo c.fonzo@inapp.gov.it
Marianna Forleo m.forleo@inapp.gov.it