FONDAZIONE BRUNO KESSLER
MAGAZINE

Governing Artificial Intelligence: the role of public policy and Europe's example

January 17, 2024

The recent surge in the artificial intelligence (AI) debate has been fueled by the release of ChatGPT and more recently the European Parliament and Council's agreement on the AI Act. This groundbreaking legislation will regulate AI use in the EU by imposing varying levels of obligations, restrictions, and prohibitions based on different risk categories. The AI discussion is no longer confined to the realms of computer science and engineering but has expanded to include philosophical, legal, economic, and sociological questions due to its wide-ranging implications.

Artificial intelligence has enormous potential to do good: initiatives such as the 'AI for Good' platform or the Think Thank 'AI4SDGs' are engaged in the study and dissemination of AI applications for social welfare. The most straightforward examples include medical applications for early disease detection, drug development, and telemedicine. AI can also be used for analysing satellite data to monitor climate change phenomena, identify malnutrition problems or track illegal fishing practices. Moreover, it finds applications in education, agriculture, finance, disability support, traffic optimization, waste management, building efficiency and more.

However, AI also carries risks that cannot be overlooked. For instance, image recognition could be used to quickly and accurately identify specific religious groups based on distinctive clothing features. In the public discourse, AI's impact on employment is a prominent concern, as it is reshaping the workforce by automating tasks, creating new roles, and transforming existing jobs. Therefore, lifelong learning as well as educational systems, social safety nets and active labour policies play a crucial role. Moreover, AI raises critical questions about inequalities, as it is not clear how these benefits and costs will be distributed across society.

AI-related risks are mainly pre-existing issues, which are exacerbated by the power of new these technologies: algorithms replicate and amplify patterns found in the data leading for instance to privacy and copyright issues, misinformation, biases and discrimination. The core problem lies precisely in their inherent strength of these technologies: the capacity for rapid and efficient large-

scale application. If the input data contain errors and biases, if the underlying purposes are unethical or the objectives not correctly specified, the computing power of the algorithms will exponentially amplify these flaws. This is particularly concerning in the context of autonomous weapons, where scalability combined with precision in target recognition poses significant dangers.

These challenges are more political than technical, requiring government intervention and public discussion. In his book "Human Compatible: AI and the Problem of control" Stuart Russel[1] delves into these ethical and governance issues. He discusses value misalignment, where machines optimise specified goals but fail to understand or align with human values. Russel uses the myth of King Midas as an analogy, who was granted his wish by Dionysus to turn everything he touched into gold. This blessing turned into a curse when Midas realized that his food also transformed into inedible gold upon contact, leading him to starvation. In a modern parallel, Russel illustrates the potential unintended consequences of AI, envisioning a scenario where a machine is programmed to mitigate the rapid ocean acidification caused by escalating carbon dioxide levels. The machine successfully develops a catalyst that rebalances the ocean's pH levels. Unfortunately, this solution comes with an unforeseen cost: a significant portion of atmospheric oxygen is depleted during the process, leaving humanity asphyxiated.

Human behaviour adds even more complexity, as machines typically rely on observed actions to predict human preferences. But behavioural sciences show that human decision-making frequently deviates from rationality, influenced by psychological and contextual factors.

Addressing AI's challenges therefore requires a multidisciplinary approach, integrating computer science, statistics and robotics with ethics, law, and social sciences, right from the systems' design phase and throughout their lifecycle.

Public policy plays a key role in this context but face the challenge ok keeping pace with rapidly advancing technology that are by nature moving much faster than legislation. In 2019, the OECD published the Principles on AI and in 2021 adopted the Recommendation for Agile Regulatory Governance to harness Innovation emphasising the necessity for flexible regulations that can adapt to technological evolution and advocating for the integration of regulatory impact assessment tools. Several countries have developed AI national strategies and the OECD AI Policy Observatory provides a comprehensive overview of these policies, highlighting an increasing governmental recognition of their role in this area.

The European Union has taken decisive action regarding AI regulation, resulting in the AI Act, agreed upon on December 9, 2023. This new regulation introduces a classification system for AI applications based on their risk levels, with corresponding obligations and restrictions for each category. AI uses classified as unacceptable risks to individual security and fundamental rights will be banned. These include biometric recognition in public places, social scoring and predictive policing systems, or emotion recognition in workplaces and educational settings – except for medical or security reasons (e.g. monitoring a pilot's fatigue levels). For AI systems identified as high-risk, the Act mandates specific compliance requirements. These include prior impact assessment, risk and quality management systems, human oversight obligations, robust data governance and proper user information. For limited or minimal risk systems, the legislation will impose only transparency obligations and encourage companies to voluntarily adopt codes of conduct.

The AI Act, expected to be enforced by the end of 2024, will standardize AI regulation across the EU's 27 member states. Although limited to EU countries, this legislation has significant extra-territorial implications as it applies to all AI systems impacting EU residents, irrespective of the

systems' development or deployment locations.

The AI Act follows other important EU digital legislation, such as the General Data Protection Regulation (GDPR), the Digital Services Act (DSA) and the Digital Markets Act (DMA) the Data Act and the CyberResilience Act. European legislation reveals a tendency to adopt a cautious approach to new technologies, favouring risk mitigation over the innovation race which characterises the national AI strategies of the US and China. Such approach, while prudent, could potentially result in a technological delay, yet still exposing Europe to potential international spill over effects. There is indeed a gap in global AI governance, only partly filled by OECD initiatives such as the definition of the AI Principles and the AI Observatory, which seek to promote international information, dialogue and collaboration.

As AI continues to revolutionize industries and streamline tasks, its ethical, legal, and social implications grow increasingly complex. Regulation and public policy are crucial for striking a balance between fostering innovation and managing risks, directing AI's potential towards socially beneficial and ethically sustainable outcomes. The AI Act is a significant step, but its effectiveness in achieving this critical balance will need to be thoroughly evaluated.

---

[1] Stuart Jonathan Russell is a professor of computer science at the University of California, Berkeley and known for his contributions to the field of artificial intelligence

PERMALINK

https://magazine.fbk.eu/en/news/governing-artificial-intelligence-the-role-of-public-policy-and-europes-example/

TAGS

- #artificial intelligence
- #Europa
- #Intelligenza artificiale
- #politiche pubbliche

AUTHORS

- Greta Sofia Lampis