9. Interview with Natali Helberger

Fabian Ferrari

Introduction

On March 13, 2024, the European Parliament approved the EU Artificial Intelligence Act (AI Act). This marks the concluding phase of an extensive process initiated in April 2021. There are still tasks pending for finalizing the text, and the law requires formal endorsement from the European Council. It is expected to take effect twenty days following its publication in the official journal, expected to be in either April or May 2024, and will be fully enforceable twenty-four months thereafter.

Professor Natali Helberger is one of the leading experts on the EU AI Act. In addition to being a distinguished university professor of law and digital technology with a special focus on AI at the University of Amsterdam and a member of the board of directors of the Institute for Information Law (IViR), Helberger co-founded the AI, Media & Democracy Lab and is a member of the Royal Netherlands Academy of Arts and Sciences. She is also the director of Public Values in the Algorithmic Society (AlgoSoc)—a ten-year research program funded by the Dutch government's Gravitation initiative that brings together scholars from the law, social sciences, computer science, and the humanities from five leading universities in the Netherlands. Helberger has also advised European institutions, such as the European Commission and the European Parliament, and worked with the Council of Europe on AI and fundamental rights, playing a vital role in shaping AI governance research in the Netherlands and the EU.

In this interview, Helberger comments on how the EU AI Act may shape the future of digital governance in the EU, fundamental rights, and public values, as well as on the role of the tech giants.

Fabian Ferrari is **FF**, Natali Helberger is **NH**.

Van Dijck, José, Karin van Es, Anne Helmond, and Fernando van der Vlist, eds. *Governing the Digital Society: Platforms, Artificial Intelligence, and Public Values*. Amsterdam: Amsterdam University Press, 2025
DOI 10.5117/9789048562718_CH09

FF: *In a recent paper published in* Computer Law & Security Review (*Helberger 2024*)*, you concluded that "in the draft AI Act, the realization of fundamental rights and European and professional values is inadvertently framed as a matter of technical formalization, standardization, and technical design choices, and platforms, technology providers and standardization bodies find themselves in the position of new arbiters of values and fundamental rights." Can you explain this argument?*

NH: The EU AI Act takes a value-driven approach, defining AI systems that pose significant risks to fundamental rights and public values as high risk and imposing a host of prevention and mitigation measures on primarily the developers of high-risk AI. For example, developers of high-risk AI must install a risk management system that allows them to continuously monitor the AI systems and identify and analyze "the known and the reasonably foreseeable risks that the high-risk AI system can pose to the health, safety or fundamental rights" (European Commission 2024). They need to have data management routines in place with the goal to examine "possible biases that are likely to affect the health and safety of persons, negatively impact fundamental rights or lead to discrimination prohibited under Union law." The systems must be designed in a way to enable human oversight to prevent or minimize "the risks to health, safety or fundamental rights" that may emerge when a high-risk system is used.

Providers of generative AI models with systemic risks need to identify and mitigate these risks to accommodate fundamental rights and public values. What is common to all these (and more requirements) is that they require that (mostly) the developers of AI solutions make an assessment under which conditions AI systems are in compliance or in conflict with fundamental rights. These systems need to be designed in a way that respects and operationalizes fundamental rights (like in the human oversight condition).

Fundamental rights are powerful commitments to core public values in our society, like the right to non-discrimination, privacy, freedom of expression, and due process. Fundamental rights, however, are also notoriously vague, and typically require interpretation in a given context, and also include the balancing of conflicting fundamental rights. So far, making this assessment has been the task of courts, government institutions, and fundamental rights experts. The core expertise of technology companies such as OpenAI, Google, Microsoft, or Meta is not fundamental rights, and in the past years we have seen rounds of further dismissals and reductions of ethics and responsible AI teams in these companies. This is why the role

of standardization bodies and the implementation acts of the European Commission will be so important. Their task will be to specify the AI Act's general references to fundamental rights and public values in the form of a series of technical standards or common specifications in the EC's implementing acts. Conformity with those harmonized standards and common specifications will create a legal presumption of conformity with the requirements of the AI Act. This is why I argue that technology providers and standardization bodies find themselves in the position of new arbiters of values and fundamental rights.

This is a break with traditional fundamental rights doctrine, according to which fundamental rights would bind in the first place public institutions and governance, which then have positive obligations to create the conditions so that citizens can also benefit from their fundamental rights in relation to private actors. Insofar the AI Act continues a trend that we could already observe in, e.g., the Digital Services Act (DSA). The degree to which the emerging digital regulatory framework is outsourcing fundamental rights to private companies is unprecedented and potentially in tension with the positive obligations of states to secure and protect our fundamental rights.

For AI governance in the EU this means that the European standardization organizations as well as the European Commission, through its implementation acts, will have a critical role in operationalizing the AI Act and fundamental rights. So, while the AI Act will set out the broader lines of AI governance in Europe, it is the technical standards and implementation acts, but also the (technical) instructions from developers to deployers and the terms and conditions of technology providers that will ultimately regulate and shape AI systems in Europe. As a result, what we are experiencing here is a technologization and bureaucratization of digital governance. In standardization bodies, traditionally technical expertise prevails. The European Commission has so far limited experience and limited competency in setting fundamental rights standards, but it has a lot of expertise in setting up processes. And a recent recruitment notice from the AI Office reads: "*Technology* specialists, hired as Contract Agents in Function Group IV, will play a pivotal role in enforcing and supervising new rules for general-purpose AI models" (European Union 2024, emphasis added). Making sure that there is sufficient fundamental rights expertise at standardization bodies, the European Commission, the AI Office, and technology companies will be a key challenge moving forward. Establishing fruitful collaborations with experts but also human rights standardization organizations such as the Council of Europe will be pivotal.

FF: *Who will benefit from the AI Act, and why will they benefit from it? Who will not benefit from the AI Act, and why will they not benefit from it?*

NH: The parties that are most likely to benefit from the AI Act are large technology providers with the necessary resources to ensure compliance and are then able to use compliance with the strict legal provisions to argue convincingly that their technologies are trustworthy. Providers of generative AI models will benefit because the majority is hardly regulated, and the few companies whose models are large enough to qualify as models with systemic risks (such as Google's Gemini and OpenAI's ChatGPT), have the resources to ensure compliance. Whether citizens will benefit depends on the operationalization of the provisions and whether the measures, technical standards, and implementation acts are successful in operationalizing fundamental rights and identifying and addressing risks to fundamental rights and public values. Note that under the AI Act, citizens have hardly any concrete rights that they can invoke (save a right to transparency and lodge a complaint in case of an infringement of the provisions of this regulation).

Who will not benefit are, for example, professional users of non-high-risk AI systems, such as media organizations that rely on third-party technology. Here, the law will hardly create any legal guarantees to ensure the safety and trustworthiness of AI systems, and it will be up to deployers (such as media organizations) to investigate and decide whether AI systems are safe to use or not. Responsible procurement will be key here.

FF: *Can the AI Act foster the competitiveness of EU companies and result in less industry dominance by American tech giants like Amazon, Google, and Microsoft?*

NH: Ensuring the competitiveness of EU companies was not an explicit goal of the AI Act, though it does seek to promote innovation and the functioning of the internal market. In terms of competitiveness, other legal frameworks are potentially more relevant, such as the Digital Markets Act or European Competition Law. There is one exception: the new rules about generative AI. Upon the successful lobby of a couple of "European champions," including the French AI company Mistral and the German Aleph Alpha, the regulation of most generative AI models is light touch, mostly transparency-related obligations, and open-source models are by and large exempted. Only the largest models, such as Google's Gemini or

OpenAI's ChatGPT, reach the threshold for further-going regulation. This creates the peculiar situation that, under the European AI Act, for the time being, only US companies have to be concerned about fundamental rights and public values—while the smaller, European generative AI models are off the hook. Meanwhile, the recent announcement of Microsoft's investment in Mistral triggered the question of how long the European champions will remain European.

FF: *What is your perspective on the legal treatment of open-source AI systems in the AI Act, specifically their subjection to exceptions* (*e.g., transparency requirements*)*?*

NH: This is a very difficult question and one that would have warranted more discussion during the making of the AI Act. Overall, I am doubtful whether open source is synonymous with transparency, and being open source does not automatically translate into being accountable or respecting fundamental rights and public values.

FF: *You have been working with future scenario writing methods as instruments to foster creative anticipatory ethical or legal reasoning by engaging diverse policy perspectives* (*Helberger 2024*)*. Is there a scenario in which the balance between public and private values is ideal?*

NH: Nice question. First of all, I do not think that the distinction between public and private values is that clear-cut. Maybe the more relevant question is: Who prioritizes whose values, and how to strike a fair balance between conflicting values? The interesting thing about the scenario method that we used is that it can be a means to engage citizens with diverse backgrounds in the question of which values they think are at stake when deploying AI and what values are important to them. Often, the value debate is led top-down by experts, companies, regulators, and civil society representatives, but the whole point with AI, and generative AI, in particular, is that the technology has left the lab for good and is everywhere in society, affecting all of us, and the values that matter to each of us. Better understanding whose values are at stake, when and how we can use more participatory approaches, should be an important element of risk assessments and doing responsible AI.

References

European Commission. 2024. Job Opportunities at the European AI Office. https://digital-strategy.ec.europa.eu/en/news/job-opportunities-european-ai-office.

European Union. 2024. Job opportunities at the European AI Office. March 7. https://digital-strategy.ec.europa.eu/en/news/job-opportunities-european-ai-office.

Helberger, Natali. 2024. "FutureNewsCorp, or How the AI Act Changed the Future of News." *Computer Law & Security Review* 52: 105915. https://doi.org/10.1016/j.clsr.2023.105915.

About the Author

Fabian Ferrari is assistant professor in cultural AI at Utrecht University.