FONDAZIONE BRUNO KESSLER
MAGAZINE

Regulating and learning about Artificial Intelligence through the challenges of deepfakes

July 8, 2024

Regulating AI is a complex and multifaceted issue that requires a balanced approach.

This was discussed on May 10 **by Silvio Ranise and Massimo Leone**, directors of Fondazione Bruno Kessler's Center for Cybersecurity and Center for Religious Studies respectively. You can view and listen to the full meeting here.

Although the main focus of recent legislation in Europe has been on funding AI research and development, this equally important issue is **how to control and regulate the implementation of AI technologies**.
A key principle underlying this regulation is **the risk-based approach**, which recognizes the potential impact of AI on people's fundamental rights and freedoms. The impact of AI-based tools means that even a threat with a low probability of materializing can have a huge negative impact because of the scale (potentially tens or hundreds of millions of users) at which such algorithms operate.

Regulators must carefully assess the likelihood and severity of negative outcomes for different stakeholders, with the stated goal of finding an optimal balance that takes into account these different perspectives. Technically, these kinds of **optimization** problems are called **multiobjective** problems and it is well known (thanks to the theory developed by the Italian economist, sociologist and engineer Vilfredo Pareto) that they do not admit a single optimal solution but many that have the following characteristic: one cannot improve the condition of one party without worsening the one of another. Therefore, to solve the optimization problem, one must choose among the many optimal solutions the one that best balances the disadvantages of the affected parties. This choice requires a variety of considerations, including ethical, legal, economic and political ones.

A further complicating of the situation lies in the **perception in interactions between human and AI** that tends to create the perception of **algorithmic omniscience.** To avoid this effect,

more awareness should be created about the benefits and limitations of AI algorithms.

To this end, “transparency is a crucial element,” **Ranise** pointed out, “as users need to be informed when interacting with AI-generated content.”

To facilitate a **more responsible relationship with AI**, techniques are in place, such as watermarking (a mechanism to derive information about the origin of digital content, such as images, videos and documents, by embedding hidden information within them), that make it easier to identify the origin of AI-generated content. However, this is a complex challenge, as hackers can quickly adapt to technical measures taken to reduce risk. The real solution may lie in cultivating a broader cultural understanding and consensus on the appropriate uses of generative AI

. “Just as language can be used both to represent reality and to create fiction,” **Leone** added, “AI tools have the power to augment human communication in both beneficial and harmful ways.”

Regulatory sandboxes (controlled environments where supervised intermediaries and technology players can test innovative products and services for a limited period of time) already offer a promising approach, allowing for controlled experimentation and rule tuning before larger-scale implementation, helping to manage the risks that adoption of such techniques can bring. Ultimately, progress will depend on developing not only technical safeguards but also the wisdom needed to use these powerful technologies responsibly.

The need for a risk-based approach to artificial intelligence

Beyond the observations that Ranise and Leone have discussed at length, the **European Union’s focus on regulating the development and deployment of AI** is a critical step in managing the profound impact this technology can have on our society. “While the vast potential of AI promise huge benefits, we must also carefully balance these technological advances with the fundamental rights and freedoms of citizens,” **Leone** warned. “The sheer complexity and scope of AI’s influence require a nuanced, risk-based approach to the way it is to be governed.”

This risk-based framework therefore recognizes that AI is not inherently “good or bad,” but its impacts are **highly dependent on the application context**. By thoroughly understanding the multiple risks associated with the use case, it enables the controlled adoption of new and innovative technology solutions while maintaining adequate protection against misuse or unintended consequences.

The power and danger of technologiy advances

The power of language in creating “fakes” is moreover a historical fact, originating long before the rise of digital technologies. Humans have always possessed the extraordinary ability to use words to construct narratives, whether grounded in reality or not, language being “the most powerful technology we have for generating fakes,” **Leone** warned. This ability has enabled rich cultural expression through fiction, poetry and philosophy, but it has also **facilitated the spread of misinformation and propaganda**. The current cat-and-mouse game between “detection methods” and “evolving techniques” for generating so-called synthetic media is simply the latest iteration of this centuries-old dynamic. As tools for manipulating audio, video and text become more sophisticated, the challenges associated with distinguishing between authentic and fictitious content become more complex.

However, Ranise and Leone remind us that this should not mean that we are doomed to a future of ubiquitous deception. The role of communication and consensus building will be crucial in shaping perceptions and social norms around these new technologies. Just as we have collectively developed cultural practices and legal frameworks to manage the risks of language-based “fakes,” we must now do the same for their digital counterparts.

This will require a delicate balance: **embracing the creative potential of generative AI while at the same time establishing barriers to prevent its misuse.**

Regulatory scenarios, transparency standards, and educational initiatives can all play a role in cultivating this new social consensus. The key is to avoid knee-jerk reactions of **techno-utopianism or apocalyptic fear, an** almost “Manichean” distinction that, moreover, occurred often in the decades before today as well. Istead, we must carefully address the nuances, harnessing the power of technology while remaining rooted in our shared human values.

It will be critical to combine sound technology awareness and a shared understanding of the capabilities and limitations of AI, both among developers and end users. Only through this combined effort of technical rigor and cultural consensus, the speakers concluded, can we address the challenges posed by artificial intelligence in a thoughtful and balanced way. Embracing complexity, rather than surrendering to simplistic prohibitions or unbridled adoption of AI, is the path toward realizing the full potential of this transformative technology while preserving our core human values.

PERMALINK

https://magazine.fbk.eu/en/news/regolamentare-e-conoscere-lintelligenza-artificiale-attraverso-le-sfide-dei-deepfake/

TAGS

- #artificialintelligence
- #cybersecurity
- #deepfake
- #disinformation
- #fakenews
- #IA
- #religiousstudies

RELATED VIDEOS

- https://www.youtube.com/watch?v=C2quaJiT5As

AUTHORS

- Andrea Franceschini