FONDAZIONE BRUNO KESSLER
MAGAZINE

Generative AI and disinformation: advances, challenges and opportunities

March 5, 2024

Four European projects join forces against disinformation with advanced AI models. New white paper from European digital media observatory available online

Across Europe and around the world, research efforts are leading to the development of state-of-the-art AI models for detecting and analyzing online disinformation. In this context, FBK plays a lead role. Under the coordination of Riccardo Gallotti, principal investigator, the AI4TRUST project, funded by the European Union under the Horizon Europe program (HORIZON-CL4-2021-HUMAN-01-27 AI to fight disinformation) was launched a year ago and over the next two years aims to build a platform that will combine the inputs of artificial intelligence and expert fact-checkers.

The purpose of the system will be to monitor various social media and information sources in near real-time, using state-of-the-art AI algorithms to analyze text, audio, and video contents in eight languages. This tool will allow content with a high risk of misinformation to be selected so that it can be flagged for review by professional fact-checkers, whose input will provide further feedback for the continuous improvement of the algorithms used. Reports tailored to the needs of media workers will also be developed, with the aim of providing reliable information to prevent the uncontrolled spread of disinformation.

The work provides for collaboration with other projects that, with different focuses, skills and technologies, are tackling the same problem. The sharing between **AI4Trust** and three other projects, **AI4Media**, **vera.ai** and **Titan**, also funded by the Horizon Europe program, resulted in a publication on generative AI and disinformation. This is an agile document that in less than forty pages analyzes the role of AI in the creation of disinformation, the detection technologies that can be employed, and the ethical and legal challenges it presents.

The white paper, entitled “Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities“, to which Riccardo Gallotti also contributed was published in February 2024 and is available online at the European Digital Media Observatory website.

The text begins with a systematic classification of synthetically generated disinformation, describing its types (AI-generated text, audio, images, and video), prevalence, and impact on elections. It goes on to discuss recent advances in the detection of synthetically generated disinformation.

It presents selected AI-based tools to assist professional fact checkers and citizens.
In the sequel, it focuses on emerging ethical and legal issues at the intersection of disinformation and generative AI, in particular examining the concept of data and information pollution, the overload of which poses a global risk, and in this regard this year represents an important test case given that about half of the world's population, more than 4 billion people, will be called to the polls. Finally, in addition to providing a very rich bibliography on the topic for further study, it describes the challenges to be met:

- Generative AI, hallucinations, and the quality of data trained/provided by large language models (LLM)
- Overcoming citizens' unfounded trust in AI
- Developing new tools to detect AI-generated content – Beyond English: new multilingual detection tools are needed – Access to data for researchers – Scarcity of research funding
- Beyond English: New multilingual detection tools are needed
- Data access for researchers
- Scarcity of research funds

The study highlights the centrality of research related to AI-generated images and videos, including coordinated campaigns and misinformation generated by ChatGPT or other similar tools. To make the best use of limited data and available funds, researchers rather than being hampered by current research practices (in which different groups tend to compete with each other to produce the best models and most cited publications) would have the chance to start collaborating in pursuit of the common goal of achieving rapid progress. For this to become possible for the benefit of society, funders and policymakers should provide an appropriate funding and collaboration framework that will enable long-term cross-border and inter-project collaborations. The societal and geopolitical impact of this joint and coordinated approach to countering online influence and misinformation would be very significant and is highly needed, as the stakes have never been higher in terms of maintaining electoral integrity, sustaining trust in the media and democracy, and impacting the health of citizens, to name but a few instances.

In terms of more concrete next steps, the white paper's authors specifically request from the EU:

- better mediation and broader and better access to data for the purpose of training new AI-based detection models, as well as overcoming the limitations of approaches to data access offered by social platforms expressly for the purpose of enabling researchers to apply new AI models to counter online misinformation.
- the provision of funding for the creation of comprehensive multilingual training datasets for researchers in all European countries. The creation of new models requires data labeled as "humans" to improve AI algorithms and evaluate their performance on different types of misinformation, spanning many European countries and languages. Such a joint and well-funded data creation initiative will therefore allow researchers to join forces in creating these much-needed but expensive datasets. Data that platforms already own, as it is created (but not shared!) as a side effect of content moderation efforts.

PERMALINK

https://magazine.fbk.eu/en/news/generative-ai-and-disinformation-advances-challenges-and-opportunities/

TAGS

- #audio
- #chatgpt
- #disinformazione
- #IA
- #Intelligenza artificiale
- #media
- #social media
- #societàdigitale
- #testo
- #verifica dei fatti
- #video

AUTHORS

- Giancarlo Sciascia