LLMs for Democratisation: Risks and Opportunities

Marco Lippi[1], Erik Longo[2], Simone Marinai[1] and Giuseppe Mobilio[2,*]

[1]*Department of Information Engineering, University of Florence*

[2]*Department of Legal Sciences, University of Florence*

Abstract

Generative artificial intelligence is bringing a deep revolution to many aspects of our society. Public authorities and representative assemblies are not an exception, as they will be profoundly concerned by this transformation. In this paper, we discuss opportunities and risks for the use of this novel technology in representative assemblies, showing how they can enable a process of democratization, while opening to possible threats and perils for the civil society.

Keywords

Large language models, Public authorities, Democratization

1. Introduction

Large Language Models (LLMs) are spreading around the world and have triggered a process of radical evolution in various aspects of our society, including the nature and functions of public authorities. Representative assemblies are among the bodies most likely to be affected by integration with these AI technologies. National parliaments are at the forefront of experimentation, not least because of the availability of data, resources, and expertise, as the Inter-Parliamentary Union's 2024 findings show [1]. However, it is not unrealistic to think that LLMs will become increasingly accessible to all levels of government, including regions and local authorities. In any case, the use of LLMs is bound to affect some of the core principles of the democratic system at all levels [2]. But the threats posed by the integration of LLMs into information, organizational and decision-making processes are well known and should be handled with care, ranging from hallucinations to cyber-security vulnerabilities [3, 4].

This paper focuses on a paradigmatic case that highlights the opportunities and threats posed by LLMs, namely their use to process and summarize those legal documents discussed and adopted in representative assemblies which reflect the political opinions of political representatives or groups, such as motions or parliamentary questions. The peculiarity of these documents is that they reproduce a language characterised by nuances, allusions or hyperbolic expressions that LLMs run the risk of not recognising and not adequately reproducing in their elaborations. Opportunities and potential risks will be analyzed both from a technological and a legal point of view, highlighting advantages or disadvantages to this use, and what technical and legal measures need to be taken.

2. Opportunities

In the last few years, generative Artificial Intelligence (AI) has produced a paradigm shift in the technological world. The unprecedented success of this novel technology is mostly due to a straightforward training procedure that allows to train very large neural networks grounding on the basic idea of predicting the next word in a sentence. The supervision for this task can be automatically constructed

given a text of arbitrary length, and can therefore exploit huge data collections to implement this self-supervision technique [5]. These pre-trained networks have shown surprising skills (named *emergent abilities*) in many heterogeneous tasks: from document summarization to question answering, from information retrieval to text comprehension and reasoning [6].

The spread of this kind of models has enabled the development of many applications across different domains, including the legal one. From a technological perspective, LLMs offer evident opportunities with respect to classic AI and machine learning methodologies. First of all, they allow even non-experts to utilize this technology, since LLMs can be tested in a straightforward way via *prompting* techniques formed directly in natural language, without requiring technical skills in computer science. Moreover, the emergent abilities induced by pre-training permit to develop a wide range of applications without the need to construct large training data sets – an activity that is usually very expensive in terms of needed time and resources, and that requires expert knowledge for highly specific domains [7].

In the context of democratic assemblies, from the legal point of view, this novel use of LLMs is bound to bring several benefits. First of all, LLMs can enable *greater democratic control*: citizens can have easier and more immediate access to official documents that express the political position of their representatives. Another related key advantage would be that of *reducing the distance between representatives and represented individuals*: LLMs can in fact lower the barrier of technical language that too often moves citizens away from the law [8]. By allowing the simplification of a large number of documents, LLMs could offer a valuable instrument to counter detachment and disaffection from political participation. In addition, generative AI could *promote the right to information*, by strengthening the right of citizens to be informed about facts and issues of general interest, as well as *promoting accountability*, by fostering citizens' control over politicians by giving them access to simplified information.

3. Risks

Clearly, the proliferation of this novel use of AI will also pose the problem of possible risks and threats to society. From a legal perspective, a first peril comes from the possibility of *undermining parliamentary debate*. In fact, generative AI tools may miss, and therefore reduce, language made up of subtleties, nuances, and compromises [9].

A consequence of this issue could be a *weakening of democratic control*. In Section 2 we have highlighted how AI could strengthen democratic control. However, there is also a concrete risk that citizens without technical expertise or lacking critical thinking, having little knowledge of how AI algorithms work, cannot exercise control or verify the outcome, except through direct access to the produced documents [10]. As a result, AI algorithms may generate errors and fail to faithfully reproduce political thinking without citizens being aware of it.

Another potential threat to society is the spread of misinformation or disinformation that inappropriate use of LLMs could encourage, also amplified by social media platforms, leading to *polluting public debate*. In such an arena, traditional media risk seeing their intermediary role weakened. From a technical point of view, this behaviour of LLMs could be the result of *hallucinations*, i.e., portions of generated text where the content is completely invented, without any specific ground.

This could pose a problem for the *stability of democratic institutions*, since the relationship of confidence between representative assemblies and executives depends on political judgment, which is subject to the pressure of public opinion. An untruth, if widely and convincingly disseminated, can have political consequences and undermine this relationship of confidence, the stability of institutions and their ability to take decisions. This process could even lead to *causing legal liability*: when LLMs contribute to the representation of a political actor or attribute an idea to a political force in an untruthful way, they cause unlawful damage (image damage, defamation, etc.) to individuals and society, and thus they could depict a scenario for legal liability.

Finally, again from a computer science perspective, there is the potential risk that very large models, that often show the best performance [6], will be in the hands of a few private companies, without the possibility, for the civil society, to have any guarantee of transparency or accountability.

4. Discussion

The opportunities and risks outlined above can be addressed by applying certain regulatory requirements and implementing technical solutions to maximise the benefits of using these technologies. At the legal level, the most recent innovation is the Regulation (EU) 2024/1689 "laying down harmonised rules on artificial intelligence" (so called AI Act), adopted in June 2024 and not yet fully in force.

One way to increase the reliability and credibility of content produced by LLMs is to use disclaimers visible on the masks for questioning the LLMs or watermarks in the generated content. The aim is to make it clear that the information is produced by an LLM and that political statements are not reproduced exactly. On a legal level, the AI Act imposes conditions in this regard by requiring providers of AI systems, including general-purpose AI systems, to mark output "in a machine-readable format and detectable as artificially generated or manipulated" (Art. 50 of AI Act). This is also an active area of research from a computer science point of view [11, 12].

Another solution is to cite the sources from which LLMs derive the processed or summarised information. This allows citizens to check whether a political position is being distorted or whether its meaning is being faithfully reproduced. Some provisions of the AI Act go in this direction by classifying LLMs used for this purpose as "high risk", because they are "intended to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda" (Annex III of the AI Act). As a result, Article 14 imposes obligations of "human oversight", which allows, for example, the output of the AI system to be correctly interpreted in order to prevent or minimise risks to fundamental rights, such as those related to information and political participation. From this point of view, we cannot hide the risks of awareness, i.e., of uncritical reliance on AI systems and the temptation to always trust the results offered by the technologies.

The AI Act also imposes general transparency requirements on the functioning of algorithms, so that their reliability can be assessed. Art. 14 requires AI systems to be "sufficiently transparent to enable deployers to interpret a system's output and use it appropriately". In this sense, the instructions for the use of the LLM should include "the characteristics, capabilities and limitations of performance", and, for example, "where applicable, the technical capabilities and characteristics of the high-risk AI system to provide information that is relevant to explain its output". This calls for innovative AI methods that are either interpretable-by-design, or that explain the predictions and decisions of a "black-box" model, following the paradigm of eXplainable AI (XAI) [13].

From a technical point of view, there are a number of solutions that can be used to improve the reliability of the information produced, also with the aim of avoiding hallucinations, citing sources and enhancing the transparency of LLMs. Retrieval-Augmented Generation (RAG) is one of the most widely employed techniques developed with this goal [14]. RAG consists in a system architecture designed to incorporate an external knowledge base (such as a database, or a memory) within the LLM architecture. By leveraging this additional amount of knowledge, prompts can be enhanced with more information related to the context. This mechanism is shown to greatly reduce hallucinations, and to improve the reliability and accountability of LLMs with respect to users, as they become able to cite the sources that support their claims and arguments [15]. In the case of representative assemblies, the external knowledge base could contain selected legal documents, such as motions or parliamentary questions, classified according to certain parameters, such as type, date of approval, context of discussion, which increase reliability. The RAG approach is often coupled with specific categories of prompts, such as the Chain-of-Thought [16], where the LLM is specifically asked to produce an answer that contains a step-by-step illustration of the reasoning path leading to the output. Recent models are also specifically designed to perform reasoning when answering a given query [17]. This aspect could strongly improve the interpretability of answers belonging to a legal context for citizens.

5. Conclusions

Generative AI tools will affect the way in which citizens perceive the actions of public authorities and representative assemblies. The proliferation of these tools will also create a new type of intermediary, different from traditional media professionals or outlets. The management and verification of the information produced by LLMs will require not only information experts, but also IT experts who can verify the reliability of the machines that process the information. At the same time, there is a growing need for legislation to clarify the legal responsibilities arising from the use of LLMs. This need arises both from the need to defend the image of individuals, such as parliamentarians, and from the need to ensure the reliability of information and the stability of democratic systems. This is the challenge that the above-mentioned AI Act, which in many parts has yet to come into force, seeks to address.

Acknowledgments

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] Inter-Parliamentary Union, World e-parliament report, 2024. URL: https://www.ipu.org/resources/publications/reports/2024-10/world-e-parliament-report-2024.

[2] P. J. Pesch, Potentials and challenges of large language models (LLMs) in the context of administrative decision-making, European Journal of Risk Regulation (2025) 1–20.

[3] L. Weidinger, J. Mellor, M. Rauh, C. Griffin, J. Uesato, P.-S. Huang, M. Cheng, M. Glaese, B. Balle, A. Kasirzadeh, et al., Ethical and social risks of harm from language models, arXiv preprint arXiv:2112.04359 (2021).

[4] L. Weidinger, J. Uesato, M. Rauh, C. Griffin, P.-S. Huang, J. Mellor, A. Glaese, M. Cheng, B. Balle, A. Kasirzadeh, et al., Taxonomy of risks posed by language models, in: Proceedings of the 2022 ACM conference on fairness, accountability, and transparency, 2022, pp. 214–229.

[5] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al., A survey of large language models, arXiv preprint arXiv:2303.18223 1 (2023).

[6] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, et al., Emergent abilities of large language models, arXiv preprint arXiv:2206.07682 (2022).

[7] J. Breton, M. M. Billami, M. Chevalier, H. T. Nguyen, K. Satoh, C. Trojahn, M. M. Zin, Leveraging LLMs for legal terms extraction with limited annotated data, Artificial Intelligence and Law (2025) 1–27.

[8] P. Pałka, F. Lagioia, R. Liepina, M. Lippi, G. Sartor, Make privacy policies longer and appoint LLM readers, Artificial Intelligence and Law (2025).

[9] K. Palonen, Making sense of parliamentary jargon: The need for parliamentary literacy, Rhetoric, Politics and Conceptual Change (2011) 13–25.

[10] M. Gerlich, AI tools in society: Impacts on cognitive offloading and the future of critical thinking, Societies 15 (2025) 6.

[11] Q. Pang, S. Hu, W. Zheng, V. Smith, Attacking LLM watermarks by exploiting their strengths, in: ICLR 2024 Workshop on Secure and Trustworthy Large Language Models, 2024.

[12] R. Zhang, S. S. Hussain, P. Neekhara, F. Koushanfar, {REMARK-LLM}: A robust and efficient water-marking framework for generative large language models, in: 33rd USENIX Security Symposium (USENIX Security 24), 2024, pp. 1813–1830.

[13] D. Minh, H. X. Wang, Y. F. Li, T. N. Nguyen, Explainable artificial intelligence: a comprehensive review, Artificial Intelligence Review (2022) 1–66.

[14] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, Advances in neural information processing systems 33 (2020) 9459–9474.

[15] K. Shuster, S. Poff, M. Chen, D. Kiela, J. Weston, Retrieval augmentation reduces hallucination in conversation, arXiv preprint arXiv:2104.07567 (2021).

[16] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al., Chain-of-thought prompting elicits reasoning in large language models, Advances in neural information processing systems 35 (2022) 24824–24837.

[17] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, X. Bi, et al., Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, arXiv preprint arXiv:2501.12948 (2025).