On the use of large language models in the water domain

Navigating the Scylla of naïve techno-optimism and the Charybdis of technology denial

Doorn, Neelke

DOI
10.1016/j.scitotenv.2023.164639

Publication date
2023

Document Version
Final published version

Published in
Science of the Total Environment

Science of the Total Environment 892 (2023) 164639

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Reply

On the use of large language models in the water domain: Navigating the Scylla of naïve techno-optimism and the Charybdis of technology denial

Dear Editor,

I would like to thank Dr. Ray for his Letter to the Editor (Ray, 2023) responding to my article "Artificial intelligence in the water domain: Opportunities for responsible use" (Doorn, 2021).

In his letter, Dr. Ray seeks to reimagine how AI, in the form of advanced language models such as OpenAI's ChatGPT, could be more effectively harnessed to address some of the issues articulated in my article. He focuses on three issues in particular: the potential of ChatGPT in revolutionizing the concepts of Theory-Guided Data Science (TGDS), ChatGPT's potential in the democratization of decision-making, and ChatGPT's potential in establishing a robust communication bridge between diverse professionals working in the water sector. While I largely agree with Dr. Ray's analysis of gaps and work that still needs to be done, I do not share his optimism about ChatGPT and Large Language Models (LLMs) more generally. Since ChatGPT is just one of the many LLMs currently being developed, I will use the term LLM and not ChatGPT to respond to Dr. Ray's points.

Before doing so, it is important to realize how LLMs work. In very simple terms, an LLM is a language model consisting of a neural network trained on huge datasets of unlabeled text. Using a "next-word prediction paradigm" these models can generate human-like responses to concrete prompts (Bubeck et al., 2023). However, LLMs are also prone to generating factually nonsensical or incorrect text. This "hallucination" (Ji et al., 2023) happens because the data on which these models are trained are often incomplete or contradictory.

Let us now look at the three points mentioned by Dr. Ray, to start with TGDS. Dr. Ray rightly stresses the inherent complexities involved with the integration of theoretical understandings of natural systems and data-driven models. True as that may be, it is good to carefully consider the rationale behind TGDS, which is to combine the strengths of data-driven methods with those of physics-based solutions, thereby avoiding some of the shortcomings one would have if one would rely solely on either data-driven approaches or theory-based approaches. It is not entirely clear how Dr. Ray sees the role of LLMs in TGDS. If his idea is to delegate all the physics theory to LLMs, the approach might fall into the trap of purely data-driven models that are black boxes with no physical meaning and little explanatory power. To put it in more blunt terms, the very reason to bring in physics is to avoid the data-driven models providing meaningless output, and this role of "safeguarding" the physics theory is exactly what hallucinating LLMs cannot perform. However, if Dr. Ray's suggestion is to use LLM as a source of inspiration to complement "traditional" physics-based approaches, LLMs may indeed have something to offer to TGDS.

The second suggestion made by Dr. Ray is to use LLMs to use the "natural language processing capabilities" of LLMs to facilitate participatory approaches. While it may be interesting to explore this use of LLMs further in a protected space, it is questionable whether it will enhance the fairness and legitimacy of policies and strategies formulated, as Dr. Ray suggests. It is now quite well-known that LLMs may be quite biased (Ferrara, 2023) and their use may in fact reproduce biases rather than making the decision-making processes more fair and legitimate. Likewise, I share Dr. Ray's view that, given the dynamic nature of our society and our environmental challenges, we cannot uncritically rely on a fixed set of predetermined values, something I stressed in my original contribution as well, but neither can we uncritically take the output of LLMs as the final answer to what the relevant values are. Especially when dealing with changing values, data-driven methods may be too conservative in the sense of reproducing existing values (and power relations) and not capturing value dynamism and emerging values.

Lastly, and in response to my plea to strengthen exchange and collaboration between water professionals and data analists, Dr. Ray suggests to use LLMs to facilitate effective communication between these groups. He refers to LLMs' "proficiency in natural language understanding and generation". Again, I find this an original suggestion and it would be interesting to see whether LLM-based tools can be used as an interface or intermediary between people who talk different "disciplinary" languages. However, I strongly disagree with Dr. Ray's reference to LLMs' "understanding of natural language". Admittedly, LLMs can produce human-like text. However, to refer to this capacity as "understanding" seems to overstate the workings of LLMs. LLMs do not comprehend text in the sense that humans do. Their output is ultimately the result of heavy statistical analyses and pattern recognition (Masters, 2023). That is exactly why LLMs sometimes run adrift when they lack the relevant data to perform these analyses. In those situations, the LLMs do not "understand" that their output is best described as hallucinatory.

Again, I would like to thank Dr. Ray for his original and thought-provoking response to my article and I hope that his Letter with my response will spur further discussion.

Declaration of competing interest

The author declares no competing interest.

Acknowledgements

This research was supported by a grant from the Dutch National Research Council NWO (grant no. VI.Vidi.195.119).

References

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M.T., Zhang, Y., 2023. Sparks of Artificial General Intelligence: Early Experiments With GPT-4. https://doi.org/10.48550/arXiv.2303.12712.

http://dx.doi.org/10.1016/j.scitotenv.2023.164639

Available online 7 June 2023

Doorn, N., 2021. Artificial intelligence in the water domain: opportunities for responsible use. Sci. Total Environ. 755, 142561. https://doi.org/10.1016/j.scitotenv.2020.142561.
Ferrara, E., 2023. Should ChatGPT Be Biased? Challenges and Risks of Bias in Large Language Models. https://doi.org/10.48550/arXiv.2304.03738.
Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., Fung, P., 2023. Survey of hallucination in natural language generation. ACM Comput. Surv. 55. https://doi.org/10.1145/3571730.
Masters, K., 2023. Medical Teacher's first ChatGPT's referencing hallucinations: lessons for editors, reviewers, and teachers. Med. Teach. 1–3. https://doi.org/10.1080/0142159X.2023.2208731.
Ray, P.P., 2023. Refining the application of artificial intelligence in the water domain: exploring the potential of ChatGPT. Sci. Total Environ.

Neelke Doorn
Delft University of Technology, Department of Technology, Policy and Management - Values, Technology and Innovation, PO Box 5015, 2600 GA Delft, the Netherlands
E-mail address: N.Doorn@tudelft.nl.