frontiers in Water

TYPE Perspective
PUBLISHED 26 February 2026
DOI 10.3389/frwa.2026.1749745

OPEN ACCESS

EDITED BY
Mariana Madruga de Brito,
Helmholtz Association of German Research Centres (HZ), Germany

REVIEWED BY
Adeyemi Oludapo Olusola,
York University, Canada
Bruce Lankford,
University of East Anglia, United Kingdom

*CORRESPONDENCE
Fredrik Schück
✉ fredrik.schuck@smhi.se

RECEIVED 19 November 2025
REVISED 10 February 2026
ACCEPTED 10 February 2026
PUBLISHED 26 February 2026

CITATION
Schück F (2026) A perspective on using large language models for human data in human-water research: Why we should be cautious? *Front. Water* 8:1749745.
doi: 10.3389/frwa.2026.1749745

A perspective on using large language models for human data in human-water research: Why we should be cautious?

Fredrik Schück [1,2*]

[1]Department of Sustainable Development, Environmental Sciences and Engineering, KTH Royal Institute of Technology, Stockholm, Sweden, [2]Hydrology Research, Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

The rise of human-water research in hydrology has increased the need for data on human behavior and decision-making. To address this demand, hydrologists have embraced methods from the social sciences, such as surveys, interviews, and agent-based modeling (ABM). However, collecting human data is expensive and time-consuming. Therefore, some social scientists have started evaluating whether Large Language Models (LLMs) could serve as a tool for generating human-like data for surveys and social simulations. This approach have the potential to transform human-water research by making access to human data faster and more affordable. Yet, human-water researchers should be cautious about adopting this method. The method has faced criticism in the social sciences, and similar caveats would apply to human-water research. While LLMs can provide responses based on different demographic personas, they fail to replicate human complexity and diversity. They also pose challenges to scientific rigor due to limited transparency and the risks of hallucinations. Most importantly, LLM-generated data fails to accurately represent marginalized groups, and undermines efforts to make human–water research more participatory, inclusive, and transformative.

KEYWORDS
coupled human-water systems, generative AI, human data, large language models, sociohydrology, water management

1 Introduction

Human-water system research has emerged as an important interdisciplinary lens to understand how humans interact with water systems in the Anthropocene. Examples of these interdisciplinary research efforts include sociohydrology, social-ecological systems (SES), and the International Association of Hydrological Sciences (IAHS) scientific decades' Panta Rhei and HELPING (Gain et al., 2021; Kreibich et al., 2025; Sivapalan et al., 2012; Arheimer et al., 2024). To explore these new perspectives, researchers need to collect human data, such as data on human behaviors, attitudes, and decisions, in addition to environmental data (Blair and Buytaert, 2016). However, this data collection is challenging as it requires considerable time, resources, and knowledge of appropriate data-collection methods (Pande and Sivapalan, 2017; Salganik, 2019; Schwitter et al., 2022).

Human-water research has previously adopted methods from the social sciences, including surveys, text mining, and agent-based modeling (ABM) (Bubeck and Thieken, 2018; Haer et al., 2020; Madruga de Brito et al., 2020; Di Baldassarre et al., 2015) to address this new data need. While these methods have enhanced our understanding of human-water interactions (Kreibich et al., 2025), the collection is challenging, especially in the last decades, due to a decline in response rate for surveys (Jabkowski and Cichocki, 2025) or a decreased ability to collect social media data (Murthy, 2024; Trezza, 2023). To address these data collection challenges, Large Language Models (LLMs) have recently received attention in social sciences as an alternative tool, capable of producing synthetic human data to generate new datasets or fill gaps in existing ones (Kozlowski and Evans, 2025; Grossmann et al., 2023). Since OpenAI released ChatGPT in 2022, LLMs have become more accessible and have changed how people, including researchers, view the capabilities of AI (Zhang et al., 2025).

In hydrology and human-water research, the adoption of LLMs remains limited but growing (Ma et al., 2025). Novel approaches include a hydrology-informed chatbot, WaterGPT (Ren et al., 2024), or the extraction of hydrological data from a vast collection of research studies (Miao et al., 2024). However, only three studies have applied LLMs in human-water research (Kadiyala et al., 2025; Batista et al., 2025; Zhu et al., 2024), highlighting a research topic in its infancy. Even so, we should be cautious about adopting LLMs in human-water research, as the application of LLM-generated human data is currently debated in the social sciences due to challenges with scientific rigor and ethics, including misinformation, replicability, and hidden biases (Rossi et al., 2024; Weidinger et al., 2022). This critique is also relevant for hydrology, especially since hydrologists generally lack formal training in studying human behavior and attitudes.

In this article, I argue that although LLMs offer potential for addressing data collection challenges in human-water research, we should be careful to incorporate them. In the following sections, I will introduce LLMs, review their emerging incorporation in social sciences and hydrology, and discuss their risks and possible paths forward.

2 Large language models in science

2.1 Brief introduction to large language models

Large language models are deep neural networks trained on vast amounts of data to interpret and generate text. The training data consists of text that is converted into tokens, e.g., words and phrases, which are then transformed into numerical presentations, called word embeddings. Word embedding is a transformation of a word to a vector so that related words are close in the vector space. The tokenized text is used to train an unsupervised transformer model that consists of either an encoder, a decoder, or both (Vaswani et al., 2017).

Earlier language models often used the encoder BERT for natural language processing (NLP) tasks while more recent models, including Gemma and GPT-4, use a decoder model structure, commonly a generative pre-trained transformer model (GPT) (Devlin et al., 2019; Wang et al., 2024). Both approaches utilize an attention mechanism that determines the importance of each word in its textual context (Vaswani et al., 2017; Bahdanau et al., 2016). Hence, LLMs combine linguistic and computational methods to produce human-like text by predicting the most probable sequence of words in response to a given prompt (Rossi et al., 2024; Raiaan et al., 2024).

The vast scale of training data is central to LLMs' ability to generate human-like text. To achieve this, LLM companies obtain text from the internet, including books, and scientific and code databases (Raiaan et al., 2024). However, the training data and its sources are rarely disclosed, and concerns have been raised that LLM companies disregard copyrights and licenses (Bail, 2024; Stokel-Walker and Van Noorden, 2023). The data are used to train models containing billions of parameters, a process that requires extensive computational resources and time (Widder et al., 2024). Some models, such as GPT-4, are fine-tuned to reduce undesirable behavior, like hate speech, and promote more polite outputs (Bail, 2024; Törnberg et al., 2023). As with the data, the fine-tuning process is often undisclosed, and thus most LLMs are black-box text predictors (Widder et al., 2024).

2.2 Large language models in social sciences

LLMs have gained traction for their potential to transform sociological methods and research (Grossmann et al., 2023; Bail, 2024; Ziems et al., 2024). Grossmann et al. (2023) argue that the LLMs' ability to mimic human behavior could redefine social science, reducing the need for using other methods such as crowdsourcing. Instead, they envision that LLMs can generate human-like data without involving participants and thus provide what Horton (2023) describes as *homo silicus*, a computational model of humans. The researcher in favor of this argues that it is a scalable, fast, and cost-effective method that also protects human participants (Agnew et al., 2024).

To explore LLMs as stand-ins for humans, sociologists have investigated how well LLMs can replicate human behavior. For instance, Dillion et al. (2023) assessed GPT-3.5 on moral scenarios and found its judgments aligned closely with human responses. Horton found that LLMs generally reproduced behavior similar to humans in four socio-economic experiments, and Mei et al. (2024) explored LLMs' performance on the "Big Five" personality traits by employing six psychological games, including prisoner's dilemma. Similarly, Aher et al. (2023) introduced the method "*Turing Experiments*" to evaluate whether GPT models could behave indistinguishably from humans in four physiological experiments, including Milgram's (1963) shock experiment. However, because GPT is trained on scientific data, the model may already be familiar with these experiments. To address this, Aher et al. altered the experiments' description, but it is uncertain if they succeeded in making the LLM unbiased to past experiment results.

In many studies, LLMs are conditioned with personalities and demographic data to alter the outputs and explore if they can take the role of a human. The traits can be gender (Mei et al., 2024; Aher et al., 2023), surnames representing different ethnic groups (Aher et al., 2023), or political affiliation (Horton, 2023). Such conditioning could be useful in applications in human-water research. For example, an LLM could model flood adaptation strategies for different demographic groups or explore the safe-development paradox by simulating risk perception. However, LLMs are trained on text and not behavior, and although behavior might be implicit, research shows LLMs' limitations in understanding basic human reasoning (Cui et al., 2023; Stevenson et al., 2025). This makes it difficult to interpret how well *homo silicus* mimic real humans.

Other studies have scaled the human-conditioning approach by incorporating LLMs in survey studies to generate synthetic responses (Argyle et al., 2023; Bisbee et al., 2023; Hämäläinen et al., 2023; Kim and Lee, 2024). For instance, Argyle examined LLMs' ability to generate opinions about political parties and voting behavior in the US when conditioning the model with demographic backstories from the American National Election Studies (ANES). They showed that LLMs' answers were similar to human respondents and argue that LLMs, "*when properly conditioned*" can produce human-like opinions along fine-grained demographic axes.

As surveys are an important tool in human-water research (Kreibich et al., 2025), for example, to study sociohydrological concepts such as the near-miss effect (Bogani et al., 2023) or flood recovery (Bubeck and Thieken, 2018), leveraging LLMs could help with data collection. However, researchers have shown that while LLMs can reproduce average answers from specific demographic groups, their ability to capture the variance and complexity of human opinions is limited (Bisbee et al., 2023; Boelaert et al., 2025), and some researchers report misrepresentation of representation of demographic groups (Wang et al., 2025; Zhou et al., 2025). This bias is unevenly distributed between demographic groups and could be worse for marginalized groups. Wang et al. (2025) found that GPT, to a larger extent, misrepresented marginalized groups, such as Black and non-binary people. They hypothesized that out-of-group populations in the training data are influenced by political discourse. LLMs also perform worse at generating non-US synthetic opinions (Abeliuk et al., 2025), probably due to the overrepresentation of English and US-centric text in the training data. These limitations pose a risk in human-water research, as individuals subjected to floods and droughts can be marginalized and may have limited influence on the data generated about them (Weidinger et al., 2022; Savelli et al., 2021).

Another important tool for social science and coupled human-water system modeling is ABMs, where humans are simulated as computational agents with their behavior governed by decision rules (Wooldridge, 2009). In earlier ABMs, the agents were often controlled by simplistic rules, but efforts have been made to create more elaborate agents to mimic the complexity of society, for example, reinforcement learning (Hung et al., 2022) and bounded rationality (Taberna et al., 2023; Wens et al., 2019). Despite this, the increased complexity of ABMs has not always contributed to explanatory power, as modeling human behavior is challenging (Sun et al., 2016), and data scarcity on multi-agent interactions limits modeling of the human-water system (Schück et al., 2025).

In light of this, researchers have noted that incorporating LLMs into ABMs could be a way forward, as it would be able to generate "*social interactions of individuals with specific characteristics and beliefs*" (Grossmann et al., 2023). So far, LLMs in ABMs have been used to explore echo chambers (Törnberg et al., 2023), human-like interactions (Kaiya et al., 2023; Park et al., 2023), and people's attitudes to gender discrimination and nuclear energy (Gao et al., 2023). Törnberg et al. (2023) used ANES data to create LLM agents with demographic personas and implemented them in an ABM that represented a social media platform where agents could share, comment, and like news articles. The authors examined how political affiliations and social media bubbles fostered a toxic environment. The integrated LLM-ABM approach allowed for controlling the platform's feed algorithms without requiring collaboration with social media companies, and it avoided exposing humans to toxic environments. However, the setup couldn't be validated, and the influence of LLMs' politeness bias was not evaluated.

2.3 Large language models in hydrology and human-water research

In hydrology, LLMs have had a limited impact so far. A few studies have discussed the implications of LLMs in hydrological research and have mainly focused on the coding and writing aid abilities of LLMs (Foroumandi et al., 2023; Halloran et al., 2023; Irvine et al., 2023) and their ability to describe hydrological concepts (Ma et al., 2025; Halloran et al., 2023; Kadiyala et al., 2024). Both Ren et al. (2024) and Xu et al. (2025) expanded this by respectively developing a hydrological-informed LLM, called WaterGPT, by finetuning an existing GPT model with hydrological data. Other hydrological studies used LLMs to extract basin areas from abstracts in their review (Miao et al., 2024) or integrated ChatGPT as a user interface for an AI-driven flood detection system (Kumbam and Vejre, 2024).

LLMs have been adopted to a more limited extent in human-water research, yet NLP has been used before, as several studies have used it to extract data from newspapers and social media (Madruga de Brito et al., 2020; Veigel et al., 2025). However, in the last 2 years, three studies have been published that incorporate LLMs in human-water research, where one study utilized LLMs to generate human-like data (Kadiyala et al., 2025). The other two papers either used LLM to perform sentiment analysis on meeting notes in a drought-affected community (Batista et al., 2025) or developed an integrated GIS-LLM chatbot to provide personalized flood-related information to increase people's flood risk perception (Zhu et al., 2024). Kadiyala et al. (2025) developed an ABM where LLMs were used to create agents with demographic personas from traits such as age, health, and cultural background. These LLM agent groups were then tasked with allocating a budget for different flood mitigation projects. With the LLM responses, the authors' could explore how the constellation of personas influenced the allocation. Similarly to the sociology articles, the authors noted that LLMs failed to provided variance within and between different demographic personas but note that LLM simulations

> *"can guide urban planning, infrastructure development, and disaster risk reduction by providing data-driven insights tailored to specific community requirements"*.

Despite the limited implementation of LLMs thus far, hydrologist researchers envision that LLMs will radically change hydrological modeling with automation and knowledge integration that was previously impossible (Ma et al., 2025). Human-water research might follow this trend to incorporate LLMs similarly to how it's happening in the social sciences.

3 Discussion

This emerging use of LLMs could enable simple implementations of human data in hydrology, potentially reducing expensive, time-consuming, and complicated data collection, preparation, and integration. However, I believe that it poses significant risks of producing believable yet flawed and unethical research. Specifically, the application of LLMs in human-water research runs the risk of increasing the gap between research and vulnerable communities as public engagement decreases. In the following discussion, I will outline both the caveats and potential opportunities to use LLMs with caution.

One issue is that LLMs can hallucinate and provide incorrect information. This issue has been met with the argument that all models are wrong, but some are useful and that model output from any model has to be validated (Rossi et al., 2024; Horton, 2023). However, understanding when the model is incorrect can be difficult, as there is no indication of uncertainty in the output. Spitale et al. (2023) called this combination of certainty and incorrectness *compelling disinformation* and showed that GPT-3 is better at creating believable disinformation than humans. For instance, ChatGPT can provide sources to research papers that do not exist (Foroumandi et al., 2023). Conversely, LLMs can be overly correct for the purpose of modeling humans, as they can give exact answers to numerical questions (Aher et al., 2023). This probably stems from that LLMs are not predicting human behavior, rather they predict the most likely groups of words (Dillion et al., 2023; Amirizaniani et al., 2024).

Another issue is that the LLMs exhibit biased behavior when representing humans. Some researchers see this as a possibility to steer the LLM by prompting it to mimic certain groups, genders, etc. (Kadiyala et al., 2025; Törnberg et al., 2023; Aher et al., 2023). However, models are generally politically left-leaning, altruistic, and pro-climate action and show insignificant difference when prompted to represent other groups (Mei et al., 2024; Santurkar et al., 2023). Since it also has a limited ability to represent marginalized and non-US populations (Mei et al., 2024; Wang et al., 2025; Abeliuk et al., 2025), the appropriateness of generated data from demographic personas should be questioned.

In addition to the limitation of low variance and misrepresentation of non-Americans, LLMs have a risk of generating stereotypes; for instance, LLMs have propagated gender and racial stereotypes (Weidinger et al., 2022; Kotek et al., 2023). This reproduction of disinformation and stereotypes complicates validation„ as outputs might align with our own prejudices and thus reinforce confirmation biases. This is problematic as researchers have suggested that LLMs can be used to scale research on underrepresented populations (Rossi et al., 2024). Additionally, LLMs are prone to value-lock-in; thus, the LLM represents values present in the training data that are no longer relevant, for instance, missing the change of the meaning of the word queer (Weidinger et al., 2022).

It is argued by some social scientists that LLM methods can be more ethical than traditional methods since LLMs can generate data where experiments would either be unethical to carry out on humans or for vulnerable groups where there are data gaps (Horton, 2023; Aher et al., 2023). However, this approach risks that communities that have historically been excluded will continue to be neglected and misrepresented as researchers take control of their story (Agnew et al., 2024). Thus, LLM-generated data of marginalized communities can reproduce patterns of historical colonialism (Rossi et al., 2024; Couldry and Mejias, 2019). I argue that this exclusion and misrepresentation of local communities is the opposite of the aim of human-water research, where interdisciplinary and participatory efforts are important (Nakamura et al., 2025). For instance, a goal of the IAHS scientific decade HELPING is to embrace local and indigenous knowledge (Arheimer et al., 2024). The use of LLMs to represent people, especially non-Western vulnerable communities, would be a step backwards for the human-water aim of participatory efforts and engaging local communities. Additionally, understudied human-water interactions might remain hidden when there is less interaction between researchers and communities due to higher reliance on LLM-data.

The lack of replicability and transparency in LLM-generated data poses a critical challenge for rigorous research. For instance, Bisbee et al. (2023) found that the generated personas were not reproducible between different versions of ChatGPT, even if identical prompts were used. The behavior of the model can also change in the same model version (Rossi et al., 2024). This becomes problematic as studies most often can only include model versions and prompts in their methods, as training data and model parameters are undisclosed by the LLM companies. As the training data is unknown, researchers also have difficulties assessing LLMs' biases (Bail, 2024). Thus, following the FAIR data standard (Wilkinson et al., 2016) and assessing the external validity will be challenging. On top of LLMs being potentially harmful to scientific rigor, the use raises ethical concerns of infringing copyright (Rahman and Santacana, 2023), environmental impact (Rillig et al., 2023), and exploitation of workers (Perrigo, 2023).

Another question is how thorough technical understanding of LLMs' we need. I believe that since most LLMs are opaque in terms of their biases, training data, and limitations, researchers must have sufficient knowledge to use them responsibly. This also includes ethical concerns, such as environmental impact. While a complete technical expertise may be unnecessary, we should be cautious about adopting technologies simply because they are easy to use.

How should we then use LLMs in human-water research? Arguably, LMs are powerful coding and writing aids. LLMs also have a strong capability in analyzing and summarizing vast amounts of text, though we should carefully check the outputs.

Additionally, they have been suggested for prototyping research ideas (Dillion et al., 2023; Park et al., 2023). For instance, researchers can test concepts before employing expensive data collection campaigns. However, it is uncertain how helpful it will be, as the results should always be validated with real data due to problems with hallucination and bias (Rossi et al., 2024).

Another possibility is open-source LLM projects built for research purposes, where the training data, model training, and fine-tuning processes are shared, like EleutherAI (Widder et al., 2024). Open-source models could thus provide more transparency and replicability, and new results show that the performance of open-source models is getting closer to that of proprietary LLMs (Bail, 2024). Another benefit is that research would become more independent of large tech companies. For instance, Meta's LLM model was open source but gradually became more opaque (Bail, 2024). Despite this progress and the prospect of open-source LLMs, there are still challenges to be addressed, such as the representation of groups, ethical considerations of collecting data, and environmental impact, as well as whether generating human data is epistemically sound (Rossi et al., 2024).

4 Conclusions

The use of LLMs could fundamentally transform how human-water research is conducted. They have the potential to simplify complex data collection and make it more accessible, potentially leading to new insights about human-water interactions. However, as illustrated in this article, it remains unclear how well these insights will translate into real-world applications and to what degree of certainty we can use the model outputs. From my perspective, LLMs pose more risks than benefits, particularly concerning ethics related to inclusion and representation. The HELPING decade and human-water research are just as much about understanding human-water systems as they are about engaging communities affected by climate change and hydrological hazards. Therefore, LLMs might steer us in the wrong direction by excluding those with knowledge and those who might need our help.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

FS: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Formal analysis, Investigation.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Acknowledgments

I am grateful for the valuable comments and suggestions from the two reviewers. I am also thankful to Derya Akbaba for her valuable feedback on the draft of this paper.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Abeliuk, A., Gaete, V., and Bro, N. (2025). Fairness in LLM-generated surveys [internet]. *arXiv*. Available online at: http://arxiv.org/abs/2501.15351 (Accessed November 1, 2025).

Agnew, W., Bergman, A. S., Chien, J., Díaz, M., El-Sayed, S., Pittman, J., et al. (2024). "The illusion of artificial inclusion," in *Proceedings of the CHI Conference on Human Factors in Computing Systems [Internet]*, 1–12. Available online at:

http://arxiv.org/abs/2401.08572 (Accessed November 7, 2025). doi: 10.1145/3613904.3642703

Aher, G. V., Arriaga, R. I., and Kalai, A. T. (2023). "Using large language models to simulate multiple humans and replicate human subject studies," in *Proceedings of the 40th International Conference on Machine Learning [Internet]*, eds. A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, 337–371. *PMLR (Proceedings of Machine Learning Research*). Available online at: https://proceedings.mlr.press/v202/aher23a.html (Accessed October 24, 2025).

Amirizaniani, M., Martin, E., Sivachenko, M., Mashhadi, A., and Shah, C. (2024). Do LLMs exhibit human-like reasoning? Evaluating theory of mind in LLMs for open-ended responses [internet]. *arXiv*. Available online at: http://arxiv.org/abs/2406.05659 (Accessed November 7, 2025).

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., and Wingate, D. (2023). Out of one, many: using language models to simulate human samples. *Polit. Anal.* 31, 337–351. doi: 10.1017/pan.2023.2

Arheimer, B., Cudennec, C., Castellarin, A., Grimaldi, S., Heal, K. V., Lupton, C., et al. (2024). The IAHS science for solutions decade, with Hydrology Engaging Local People IN one Global world (HELPING). *Hydrol. Sci. J.* 69, 1417–1435. doi: 10.1080/02626667.2024.2355202

Bahdanau, D., Cho, K., and Bengio, Y. (2016). Neural machine translation by jointly learning to align and translate [internet]. *arXiv*. Available online at: http://arxiv.org/abs/1409.0473 (Accessed October 24, 2025).

Bail, C. A. (2024). Can generative AI improve social science? *Proc. Natl. Acad. Sci.* 121:e2314021121. doi: 10.1073/pnas.2314021121

Batista, T. L., de Carvalho Studart, T. M., Duarte, M. G., and Souza Filho, F. D. A. D. (2025). Modeling sentiment–hydrology interaction using LLM: insights for adaptive governance in Ceará's water management. *Water*. 17:2615. doi: 10.3390/w17172615

Bisbee, J., Clinton, J., Dorff, C., Kenkel, B., and Larson, J. (2023). Synthetic replacements for human survey data? The perils of large language models [internet]. *SocArXiv*. Available online at: https://osf.io/5ecfa (Accessed October 24, 2025).

Blair, P., and Buytaert, W. (2016). Socio-hydrological modelling: a review asking 'why, what and how?' Hydrol Earth Syst *Sci*. 20, 443–478. doi: 10.5194/hess-20-443-2016

Boelaert, J., Coavoux, S., Ollion, É., Petev, I., and Präg, P. (2025). Machine bias. How do generative language models answer opinion polls? *Sociol. Methods Res.* 54, 1156–1196. doi: 10.1177/00491241251330582

Bogani, A., Faccenda, G., Riva, P., Richetin, J., Pancani, L., and Sacchi, S. (2023). The near-miss effect in flood risk estimation: a survey-based approach to model private mitigation intentions into agent-based models. *Int. J. Disaster Risk Reduct. [Internet]* 89:103629. doi: 10.1016/j.ijdrr.2023.103629

Bubeck, P., and Thieken, A. H. (2018). What helps people recover from floods? Insights from a survey among flood-affected residents in Germany. *Reg. Environ. Change* 18, 287–296. doi: 10.1007/s10113-017-1200-y

Couldry, N., and Mejias, U. A. (2019). Data colonialism: rethinking big data's relation to the contemporary subject. *Telev. New Media*. 20, 336–349. doi: 10.1177/1527476418796632

Cui, R., Lee, S., Hershcovich, D., and Søgaard, A. (2023). What does the failure to reason with 'respectively' in zero/few-shot settings tell us about language models? [Internet]. *arXiv.* Available online at: http://arxiv.org/abs/2305.19597 (Accessed October 31, 2025).

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding [internet]. *arXiv*. Available online at: http://arxiv.org/abs/1810.04805 (Accessed October 24, 2025).

Di Baldassarre, G., Viglione, A., Carr, G., Kuil, L., Yan, K., Brandimarte, L., et al. (2015). Debates—perspectives on socio-hydrology: capturing feedbacks between physical and social processes. *Water Resour. Res*. 51, 4770–4781. doi: 10.1002/2014WR016416

Dillion, D., Tandon, N., Gu, Y., and Gray, K. (2023). Can AI language models replace human participants? *Trends Cogn. Sci*. 27, 597–600. doi: 10.1016/j.tics.2023.04.008

Foroumandi, E., Moradkhani, H., Sanchez-Vila, X., Singha, K., Castelletti, A., and Destouni, G. (2023). ChatGPT in hydrology and earth sciences: opportunities, prospects, and concerns. *Water Resour. Res.* 59:*e*2023WR036288. doi: 10.1029/2023WR036288

Gain, A. K., Hossain, M. S., Benson, D., Di Baldassarre, G., Giupponi, C., and Huq, N. (2021). Social-ecological system approaches for water resources management. *Int. J. Sustain. Dev. World Ecol*. 28, 109–124. doi: 10.1080/13504509.2020.1780647

Gao, C., Lan, X., Lu, Z., Mao, J., Piao, J., Wang, H., et al. (2023). S3: social-network simulation system with large language model-empowered agents [internet]. *arXiv.* Available online at: https://arxiv.org/abs/2307.14984 (Accessed October 24, 2025). doi: 10.2139/ssrn.5604712

Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E., and Cunningham, W. A. (2023). AI and the transformation of social science research. *Science [Internet]* 380:1108. doi: 10.1126/science.adi1778

Haer, T., Husby, T. G., Botzen, W. J. W., and Aerts, J. C. J. H. (2020). The safe development paradox: an agent-based model for flood risk under climate change in the European Union. *Glob. Environ. Change Hum. Policy Dimens*. 60:102009. doi: 10.1016/j.gloenvcha.2019.102009

Halloran, L. J. S., Mhanna, S., and Brunner, P. (2023). AI tools such as ChatGPT will disrupt hydrology, too. *Hydrol. Process*. 37:e14843. doi: 10.1002/hyp.14843

Hämäläinen, P., Tavast, M., and Kunnari, A. (2023). "Evaluating large language models in generating synthetic HCI research data: a case study," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems [Internet]* (Hamburg Germany: ACM), 1–19. Available online at: https://dl.acm.org/doi/10.1145/3544548.3580688 (Accessed November 1, 2025). doi: 10.1145/3544548.3580688

Horton, J. (2023). *Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus? [Internet]*. Cambridge, MA: National Bureau of Economic Research, w31122. Report No.: w31122. Available online at: http://www.nber.org/papers/w31122.pdf (Accessed October 24, 2025). doi: 10.3386/w31122

Hung, F., Son, K., and Yang, Y. C. E. (2022). Investigating uncertainties in human adaptation and their impacts on water scarcity in the Colorado river Basin, United States. *J. Hydrol. [Internet]* 612:128015. doi: 10.1016/j.jhydrol.2022.128015

Irvine, D. J., Halloran, L. J. S., and Brunner, P. (2023). Opportunities and limitations of the ChatGPT advanced data analysis plugin for hydrological analyses. *Hydrol. Process*. 37:e15015. doi: 10.1002/hyp.15015

Jabkowski, P., and Cichocki, P. (2025). Survey response rates in European comparative surveys: a 20-year decline irrespective of sampling frames or survey modes. *Qual. Quant*. 59, 635–655. doi: 10.1007/s11135-024-01993-9

Kadiyala, L., Sajja, R., Sermet, Y., Muste, M., and Demir, I. (2025). AI-driven decision-making for water resource planning and hazard mitigation using automated multi-agents. *J. Hydroinform.* 27, 1217–1237. doi: 10.2166/hydro.2025.042

Kadiyala, L. A., Mermer, O., Samuel, D. J., Sermet, Y., and Demir, I. (2024). The implementation of multimodal large language models for hydrological applications: a comparative study of GPT-4 vision, Gemini, LLaVa, and multimodal-GPT. *Hydrology*11:148. doi: 10.3390/hydrology11090148

Kaiya, Z., Naim, M., Kondic, J., Cortes, M., Ge, J., Luo, S., et al. (2023). Lyfe agents: generative agents for low-cost real-time social interactions [internet]. *arXiv*. Available online at: http://arxiv.org/abs/2310.02172 (Accessed November 5, 2025).

Kim, J., and Lee, B. (2024). AI-augmented surveys: leveraging large language models and surveys for opinion prediction [internet]. *arXiv*. Available online at: http://arxiv.org/abs/2305.09620 (Accessed October 24, 2025).

Kotek, H., Dockum, R., and Sun, D. (2023). "Gender bias and stereotypes in large language models," in *Proceedings of The ACM Collective Intelligence Conference [Internet]* (Delft Netherlands: ACM), 12–24. Available online at: https://dl.acm.org/doi/10.1145/3582269.3615599. doi: 10.1145/3582269.3615599 (Accessed November 7, 2025).

Kozlowski, A. C., and Evans, J. (2025). Simulating subjects: the promise and peril of artificial intelligence stand-ins for social agents and interactions. *Sociol. Methods Res.* 54, 1017–1073. doi: 10.1177/00491241251337316

Kreibich, H., and Sivapalan, M., AghaKouchak A, Addor N, Aksoy H, Arheimer B, et al. (2025). Panta Rhei: a decade of progress in research on change in hydrology and society. *Hydrol. Sci. J.* 70, 1210–1236. doi: 10.1080/02626667.2025.2469762

Kumbam, P. R., and Vejre, K. M. (2024). FloodLense: a framework for ChatGPT-based real-time flood detection [internet]. *arXiv*. Available online at: http://arxiv.org/abs/2401.15501 (Accessed November 5, 2025).

Ma, Z., Peng, B., Yue, Z., Zeng, H., Pan, M., Wu, X., et al. (2025). Embracing large language model (LLM) technologies in hydrology research. *Environ. Res. Water* 1:022001. doi: 10.1088/3033-4942/addd43

Madruga de Brito, M., Kuhlicke, C., and Marx, A. (2020). Near-real-time drought impact assessment: a text mining approach on the 2018/19 drought in Germany. *Environ. Res. Lett*. 15:1040a9. doi: 10.1088/1748-9326/aba4ca

Mei, Q., Xie, Y., Yuan, W., and Jackson, M. O. (2024). A turing test of whether AI chatbots are behaviorally similar to humans. *Proc. Natl. Acad. Sci.* 121:e2313925121. doi: 10.1073/pnas.2313925121

Miao, C., Hu, J., Moradkhani, H., and Destouni, G. (2024). Hydrological research evolution: a large language model-based analysis of 310,000 studies published globally between 1980 and 2023. *Water Resour. Res.* 60:e2024WR038077. doi: 10.1029/2024WR038077

Milgram, S. (1963). Behavioral study of obedience. *J. Abnorm. Soc. Psychol.* 67, 371–378. doi: 10.1037/h0040525

Murthy, D. (2024). Sociology of Twitter/X: trends, challenges, and future research directions. *Annu. Rev. Sociol.* 50, 169–190. doi: 10.1146/annurev-soc-031021-035658

Nakamura, S., Kreibich, H., Haeffner, M., Mukherjee, J., Di Baldassarre, G., Sakamoto, M., et al. (2025). Expanding and mainstreaming sociohydrology toward transdisciplinary praxis. *Front. Water [Internet]* 7:1629905. doi: 10.3389/frwa.2025.1629905

Pande, S., and Sivapalan, M. (2017). Progress in socio-hydrology: a meta-analysis of challenges and opportunities. *WIREs Water* 4:e1193. doi: 10.1002/wat2.1193

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). "Generative agents: interactive simulacra of human behavior," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology [Internet]* (UIST'23) (New York, NY, USA: Association for Computing Machinery). doi: 10.1145/3586183.3606763

Perrigo, B. (2023). Exclusive: the $2 per hour workers who made ChatGPT safer. *TIME [Internet]*. Available online at: https://time.com/6247678/openai-chatgpt-kenya-workers/ (Accessed November 9, 2025).

Rahman, N., and Santacana, E. (2023). "Beyond fair use: legal risk evaluation for training LLMs on copyrighted text," in *ICML Workshop on Generative AI and Law* (Accessed November 9, 2025).

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., et al. (2024). A review on large language models: architectures, applications, taxonomies, open issues and challenges. *IEEE Access*. 12, 26839–26874. doi: 10.1109/ACCESS.2024.3365742

Ren, Y., Zhang, T., Dong, X., Li, W., Wang, Z., He, J., et al. (2024). WaterGPT: training a large language model to become a hydrology expert. *Water* 16:3075. doi: 10.3390/w16213075

Rillig, M. C., Ågerstrand, M., Bi, M., Gould, K. A., and Sauerland, U. (2023). Risks and benefits of large language models for the environment. *Environ. Sci. Technol*. 57, 3464–3466. doi: 10.1021/acs.est.3c01106

Rossi, L., Harrison, K., and Shklovski, I. (2024). The problems of LLM-generated data in social science research. *Sociologica* 18, 145–168. doi: 10.6092/issn.1971-8853/19576

Salganik, M. J. (2019). *Bit by Bit: Social Research in the Digital Age. First Paperback Printing*. Princeton, Oxford: Princeton University Press, 423.

Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., and Hashimoto, T. (2023). Whose opinions do language models reflect? [Internet]. *arXiv*. Available online at: http://arxiv.org/abs/2303.17548 (Accessed November 7, 2025).

Savelli, E., Rusca, M., Cloke, H., and Di Baldassarre, G. (2021). Don't blame the rain: social power and the 2015–2017 drought in Cape Town. *J. Hydrol*. 594:125953. doi: 10.1016/j.jhydrol.2020.125953

Schück, F., Arheimer, B., Mazzoleni, M., and Brandimarte, L. (2025). A systematic mapping review of hydrological hazard management in agent-based systems. *Environ. Res. Lett*. 20:113003. doi: 10.1088/1748-9326/ae0fb2

Schwitter, N., Pretari, A., Marwa, W., Lombardini, S., and Liebe, U. (2022). Big data and development sociology: an overview and application on governance and accountability through digitalization in Tanzania. *Front. Sociol. [Internet]* 7:909458. doi: 10.3389/fsoc.2022.909458

Sivapalan, M., Savenije, H. H. G., and Blöschl, G. (2012). Socio-hydrology: a new science of people and water. *Hydrol. Process*. 26, 1270–1276. doi: 10.1002/hyp.8426

Spitale, G., Biller-Andorno, N., and Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. *Sci. Adv.* 9:eadh1850. doi: 10.1126/sciadv.adh1850

Stevenson, C. E., Pafford, A., van der Maas, H. L. J., and Mitchell, M. (2025). Can large language models generalize analogy solving like children can? [Internet]. *arXiv*. Available online at: http://arxiv.org/abs/2411.02348 (Accessed October 31, 2025).

Stokel-Walker, C., and Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature* 614, 214–216. doi: 10.1038/d41586-023-00340-6

Sun, Z., Lorscheid, I., Millington, J. D., Lauf, S., Magliocca, N. R., Groeneveld, J., et al. (2016).Simple or complicated agent-based models? A complicated issue. *Environ. Model Softw*. 86, 56–67. doi: 10.1016/j.envsoft.2016.09.006

Taberna, A., Filatova, T., Hadjimichael, A., and Noll, B. (2023). Uncertainty in boundedly rational household adaptation to environmental shocks. *Proc. Natl. Acad. Sci. U. S. A. [Internet]* 120:2215675120. doi: 10.1073/pnas.2215675120

Törnberg, P., Valeeva, D., Uitermark, J., and Bail, C. (2023). Simulating social media using large language models to evaluate alternative news feed algorithms [internet]. *arXiv*. Available online at: http://arxiv.org/abs/2310.05984 (Accessed Septembet 29, 2025).

Trezza, D. (2023). To scrape or not to scrape, this is dilemma. The post-API scenario and implications on digital research. *Front. Sociol. [Internet]* 8:1145038. doi: 10.3389/fsoc.2023.1145038

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need [internet]. *arXiv*. Available online at: http://arxiv.org/abs/1706.03762 (Accessed October 24, 2025). doi: 10.65215/ysbyhc05

Veigel, N., Kreibich, H., de Bruijn, J. A., Aerts, J. C. J. H., and Cominola, A. (2025). Content analysis of multi-annual time series of flood-related Twitter (X) data. *Nat. Hazards Earth Syst. Sci*. 25, 879–891. doi: 10.5194/nhess-25-879-2025

Wang, A., Morgenstern, J., and Dickerson, J. P. (2025). Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nat. Mach. Intell*. 7, 400–411. doi: 10.1038/s42256-025-00986-z

Wang, Z., Chu, Z., Doan, T. V., Ni, S., Yang, M., and Zhang, W. (2024). History, development, and principles of large language models-an introductory survey [internet]. *arXiv*. Available online at: http://arxiv.org/abs/2402.06853 (Accessed October 24, 2025).

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P. S., Mellor, J., et al. (2022). "Taxonomy of risks posed by language models," in *2022 ACM Conference on Fairness Accountability and Transparency [Internet]* (Seoul Republic of Korea: ACM), 214–229. Available online at: https://dl.acm.org/doi/10.1145/3531146.3533088 (Accessed November 9, 2025). doi: 10.1145/3531146.3533088

Wens, M., Johnson, J., Zagaria, C., and Veldkamp, T. (2019). Integrating human behavior dynamics into drought risk assessment-a sociohydrologic, agent-based approach. *WILEY Interdiscip. Rev. WATER* 6:e1345. doi: 10.1002/wat2.1345

Widder, D. G., Whittaker, M., and West, S. M. (2024). Why 'open' AI systems are actually closed, and why this matters. *Nature* 635, 827–833. doi: 10.1038/s41586-024-08141-1

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*. 3:160018. doi: 10.1038/sdata.2016.18

Wooldridge, M. (2009). *An Introduction to Multiagent Systems*. Chichester: John Wiley Sons.

Xu, B., Li, Z., Yang, Y., Wu, G., Wang, C., Tang, X., et al. (2025). Evaluating and advancing large language models for water knowledge tasks in engineering and research. *Environ. Sci. Technol. Lett*. 12, 289–296. doi: 10.1021/acs.estlett.5c 00038

Zhang, Y., Khan, S. A., Mahmud, A., Yang, H., Lavin, A., Levin, M., et al. (2025). Exploring the role of large language models in the scientific method: from hypothesis to discovery. *NPJ Artif. Intell*. 1:14. doi: 10.1038/s44387-025-00019-5

Zhou, M., Yu, L., Geng, X., and Luo, L. (2025). ChatGPT vs social surveys: probing objective and subjective silicon population [internet]. *arXiv*. Available online at: http://arxiv.org/abs/2409.02601 (Accessed November 1, 2025).

Zhu, J., Dang, P., Cao, Y., Lai, J., Guo, Y., Wang, P., et al. (2024). A flood knowledge-constrained large language model interactable with GIS: enhancing public risk perception of floods. *Int. J. Geogr. Inf. Sci*. 38, 603–625. doi: 10.1080/13658816.2024.2306167

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., and Yang, D. (2024). Can large language models transform computational social science? *Comput. Linguist*. 50, 237–291. doi: 10.1162/coli_a_00502