Check for updates

OPEN LETTER

AI for science needs data – and here come ethical problems

[version 1; peer review: 2 approved with reservations]

Kristina Khutsishvili [1], Ronald Wilders [2]

[1]University of Cambridge Department of Engineering, Cambridge, England, UK
[2]Amsterdam Cardiovascular Sciences, Department of Medical Biology, Amsterdam, The Netherlands

v1 **First published:** 29 May 2026, **6**:183
https://doi.org/10.12688/openreseurope.23390.1
Latest published: 29 May 2026, **6**:183
https://doi.org/10.12688/openreseurope.23390.1

Abstract

The contemporary discourse on artificial intelligence (AI) in scientific research is oftentimes driven by the assumption that larger datasets yield better models, and that such models inherently improve scientific understanding or technological capacity. In practice, however, the pursuit of data – especially sensitive human data in the field of healthcare – reveals fundamental ethical and governance tensions. Below we outline three core problems that require rigorous analysis: data scarcity, legitimacy, and the semantic inflation of "AI" as a label that obscures distinct technologies and ethical concerns.

Keywords

Artificial intelligence; AI for science; healthcare data; data ethics; governance; NHS; consent

Open Peer Review

Approval Status ? ?

	1	2
version 1 29 May 2026	? view	? view

1. **Tommaso Turchi**, University of Pisa, Pisa, Italy

2. **Caspar J. Van Lissa**, Tilburg University, Tilburg, The Netherlands

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Kristina Khutsishvili (kk863@cam.ac.uk)

Author roles: Khutsishvili K: Conceptualization, Writing – Original Draft Preparation; **Wilders R**: Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070325 (Innovative Solutions Responding to the Needs of Cities & Communities [CommuniCity]). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Khutsishvili K and Wilders R. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Khutsishvili K and Wilders R. **AI for science needs data – and here come ethical problems [version 1; peer review: 2 approved with reservations]** Open Research Europe 2026, **6**:183 https://doi.org/10.12688/openreseurope.23390.1

First published: 29 May 2026, **6**:183 https://doi.org/10.12688/openreseurope.23390.1

Scientific value, scarcity, and sensitivity

A central paradox of AI in science is that its highest impact applications depend on data that are simultaneously rare and ethically sensitive.

Effective healthcare models require extensive, longitudinal population-level data to capture the diversity of clinical trajectories. The scarcity of such data drives research programmes toward repositories like the United Kingdom's National Health Service (NHS) patient records, which comprise millions of patient histories. Technical proponents argue that only national-scale datasets can provide sufficient diversity to achieve the precision that is required for reliable and meaningful forecasts, including for rare diseases and minority populations. The NHS England 'Foresight' project – involving the training of generative AI on approximately 57 million de-identified patient records – was promoted as a way to "improve patient outcomes" by anticipating clinical events before they occur, thus shifting healthcare from reactive to preventative.[1]

Notwithstanding the purported clinical value, population health data are scarce in a practical sense. Such data contain deeply personal information, gathered under expectations of clinical confidentiality and governance. The richer and more detailed the dataset, the harder it is to fully anonymise in a way that effectively eliminates the risk of re-identification. In the UK, regulatory authorities explicitly treat anonymisation not as a guarantee of acceptability, but as a risk-management judgement that depends on the complexity and context of the data, rather than as a binary status that per se resolves concerns.[2]

This combination of high utility and high ethical sensitivity means that the most scientifically valuable data also carry the greatest moral weight. This moral weight, however, does not arise from the intrinsic nature of the data alone: as Solove has argued, the regulatory focus on the nature of data itself becomes unworkable in an era of large-scale analytics, where seemingly innocuous data can generate sensitive inferences and real-world harms depending on context and use.[3] From this perspective, ethical and legal evaluation cannot rely solely on whether data are formally classified as sensitive or anonymised. Instead, it needs to assess the risk and magnitude of harm arising from specific circumstances. In high-impact scientific applications, this means that even lawfully accessed and technically anonymised data may remain ethically weighty when the potential 'downstream' consequences are substantial.

Public consent, trust, legitimacy, and institutional debate

Once data scarcity motivates access to sensitive human data, the question of legitimacy emerges. Legitimacy, in this sense, involves (a) legal compliance, (b) the meaningful consent of data subjects or justified ethical waivers, and (c) public understanding and acceptance of how data are being repurposed.

The current UK health governance frameworks do not require explicit consent for all secondary uses of health data. Instead, they rely on procedural safeguards and trust in institutional governance.

The aforementioned 'Foresight' initiative has faced public and professional pushback precisely because of perceived gaps in legitimacy.[4]

In June 2025, concerns about the Foresight AI project were raised by professional bodies representing general practitioners through the Joint GP IT Committee. The committee asked NHS England to refer the matter to the UK Information Commissioner's Office in order to clarify whether data protection and governance requirements had been met. NHS England subsequently agreed to pause the Foresight project while a review was conducted and while discussions with the Information Commissioner's Office took place.[5,6]

This episode exposes an important, if not crucial, issue that is often overlooked in technical AI debates: clinical professionals and patient advocates are not always convinced that research governance matches patient expectations. In this particular case, professional bodies representing clinicians – who, in turn, play a fiduciary role with respect to patient trust – openly questioned whether governance processes and public expectations had been met in using GP records for model training.

Furthermore, there is little evidence to suggest that there is either a broad public consensus or active public engagement regarding the use of national health records to train large generative models.

The "AI" label and ethical confusion

A further ethical problem is posed by the use of "AI" as a catch-all label. The tendency to treat "AI" as a single, monolithic technology rather than a heterogeneous set of methods obscures its distinct uses and ethical profiles. When everything

from simple rule-based decision support to generative models is labelled "AI", ethical debates rapidly become muddled, and controversies from one domain (e.g., 'hallucinations' – fabricated outputs in large language models) are imported into others (e.g., clinical prediction) without proper qualification.

This semantic inflation has two harmful effects. First, it allows everyday controversies – for instance, debates about economy, markets, geopolitics, and the so-called "AI race" – to overshadow the precise ethical analysis of scientific data governance. Second, it conflates distinct ethical issues – such as algorithmic bias and informed consent – into a single "AI ethics" discussion that can neither be effectively operationalised nor mapped to practical governance levers. The NHS guidance on AI, for example, emphasises that AI systems may use personal data for training purposes, provided that the data are minimised and the confidentiality is ensured, but it does not differentiate between ethical expectations by technical subtype or societal impact of the specific model.[7]

Thus, the project of AI for science must disentangle the science of interest from the generic "AI" label. A predictive model trained on well-curated clinical data for epidemiological insight differs significantly, both ethically and technically, from a generative model trained on uncurated internet text. Conflating these under a single banner risks obscuring the very ethical trade-offs that ought to be subject to governance.

Conclusion

Addressing the ethical problems in AI for science requires simultaneous progress on three fronts:

1. Explicit governance of data use that acknowledges scarcity and sensitivity.

   This includes transparent governance processes, a clear articulation of the lawful basis for processing personal data under the UK General Data Protection Regulation (GDPR), compliance with common law duties of confidentiality, and – if relevant – explicit statutory authorisation to set those duties aside.

2. Institutional legitimacy exceeding simple compliance – and grounded in participation.

   Professional bodies, patients, and – not less importantly – the public must be involved in shaping the use of data. Being informed afterwards is a way to diminish not only public trust in bureaucracy, but also in technology itself, and, perhaps most importantly, in science and scientific institutions.

3. Conceptual disambiguation of "AI".

   Ethical analysis should clearly differentiate between types of systems (predictive, generative, rule-based clinical decision support, etc.) and reflect on specific ethical challenges into these categories.

Only with these steps taken can AI for science justify its reliance on data that are both scientifically valuable and ethically sensitive. Without this clarity, the field risks replicating broader controversies and undermining public trust.

Disclaimer

The views expressed in this article are those of the authors. Publication in Open Research Europe does not imply endorsement of the European Commission.

Data availability

No datasets were generated or analysed during the current study. This article is based on publicly available sources, which are cited within the reference list. Therefore, there are no data to deposit in a repository and no associated DOI or accession number.

Ethics and consent

Ethical approval and consent were not required for this study. The article is based exclusively on analysis of publicly available documents and does not involve human participants, personal data, or identifiable information.

Acknowledgments

The authors acknowledge the use of AI tools for assistance with formatting and language refinement.

References

1. NHS England: **Foresight AI Case Study.** 2025.
Reference Source
2. NHS England: **Proof-of-Concept: Using Anonymous or Artificial Health Data.** 2025.
Reference Source
3. Solove DJ: **Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data.** *Northwest. Univ. Law Rev.* 2024; **118**: 1081–1138.
Reference Source
4. Wilkins A: **Concerns Raised Over AI Trained on 57 Million NHS Medical Records.** *New Sci.* 2025.
Reference Source
5. Armstrong S: **NHS England Faces Investigation Over Granting Foresight AI Access to GP Data.** *BMJ.* 2025; **389**: r1192.
Publisher Full Text
6. Royal College of General Practitioners: **RCGP Raises Concerns Over Foresight AI Through Joint GP IT Committee.** 2025.
Reference Source
7. NHS England: **Artificial Intelligence.** 2025.
Reference Source

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 09 July 2026

https://doi.org/10.21956/openreseurope.25308.r75405

© **2026 Van Lissa C.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Caspar J. Van Lissa
Tilburg University, Tilburg, The Netherlands

As this is a short letter, I will constrain myself to a short comment: I completely agree with the points raised in this letter, but they are only two pieces of the puzzle (data ethics and disambiguation of the catchall term "AI"), where many other pieces of the puzzle are highly relevant (e.g., public-private partnerships where citizens' data is used to train models that are then commercialized, potentially exploiting the very citizens whose data provided the value), and misaligned incentives (a company has an incentive to convince consumers that their proprietary "AI" is superior, even if simpler, maybe free open source models actually outperform, see Rudin 2019), et cetera.

What I'm missing in this letter is: 1) A clear statement of the goal of the letter, its scope, and intended audience
2) Appropriate *scientific* references to back the points made, not just government documents
3) More integration between the points made.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
No

Does the article adequately reference differing views and opinions?
Partly

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Yes

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: methodology, interpretable machine learning, research synthesis methods, open science

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 22 June 2026

https://doi.org/10.21956/openreseurope.25308.r75409

© 2026 Turchi T. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Tommaso Turchi
University of Pisa, Pisa, Italy

In this open letter, the authors argue that the pursuit of ever larger datasets for AI in science, and in healthcare in particular, raises ethical and governance tensions that technical debates tend to overlook. They organize the argument around three problems: that the most scientifically valuable health data are also the most ethically sensitive and the hardest to anonymize; that legitimacy, covering legal compliance, meaningful consent, and public acceptance, is often assumed rather than secured; and that the term AI is used so broadly that it obscures the distinct ethical profiles of different methods. They close with recommendations on data governance, participatory legitimacy, and conceptual disambiguation of AI.

This is a timely and clearly written contribution, and the Foresight case is a well-chosen, current illustration of the legitimacy gap the authors describe. The piece is accessible and the call to stop treating AI as a single monolithic technology is well taken. The one point I would consider essential to address is the framing: the title and abstract promise an argument about AI for science in general, whereas the letter is really about health and clinical data. My remaining comments are intended to strengthen the piece rather than to question its value.

- The title and abstract frame the argument as being about AI for science in general, but the analysis is almost entirely about healthcare and the UK clinical data context. Fields such as astronomy, materials science, or climate modelling rely heavily on AI without involving sensitive

human data, and the three problems apply to them quite differently. I would suggest to either narrow the title and framing to health or clinical data, or to add a short passage justifying the generalization.

- The position of technical proponents is presented briefly but engaged with lightly, and the overall stance is predominantly critical. I would suggest giving more room to the case for secondary use under current frameworks, for instance the public-interest justification, existing mechanisms such as the national data opt-out in England, and the practical difficulty of individual consent at population scale, before explaining why the authors find these insufficient.

- A few central claims are stated without citation, notably that there is little public consensus or engagement around training generative models on health records, and that the semantic inflation of AI produces specific harms. I would suggest supporting these where evidence exists, or to mark them clearly as conjecture where it does not.

- The letter treats data as a quantity to be accessed, but what counts as data and which features are selected is itself a value-laden design decision that can introduce bias and can undermine anonymization, since combinations of innocuous features can act as quasi-identifiers. I wonder whether this dimension should be also discussed, namely that feature construction and selection is an ethical decision point in its own right, not only the question of access.

- The letter diagnoses the risks of centralizing sensitive data but does not mention the approaches developed to reduce the need to do so, most notably federated learning, where a model is trained across institutions without pooling the records. I would suggest to at least mention federated learning, and optionally related approaches such as synthetic data or trusted research environments.

- It would help to situate the argument in the European regulatory landscape rather than the UK context alone, in particular the European Health Data Space, which directly governs the secondary use of health data for research.

- The recommendations call both for public involvement in shaping data use and for a clearer disambiguation of AI into its distinct types, but do not connect the two. Meaningful participation presupposes that non-experts can grasp the distinctions the authors are drawing, so empowering people to take part requires explaining these concepts to them in accessible terms. I would suggest making this link explicit and to add public-facing explanation as a concrete next step, which would also make the recommendations more actionable.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)

Yes

Does the article adequately reference differing views and opinions?

Partly

Are all factual statements correct, and are statements and arguments made adequately

supported by citations?
Partly

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Human-Centred AI

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.