The American Journal of Bioethics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uajb20

# "Large Language Models" Do Much More than Just Language: Some Bioethical Implications of Multi-Modal AI

Joshua August Skorburg, Kristina L. Kupferschmidt & Graham W. Taylor

**To cite this article:** Joshua August Skorburg, Kristina L. Kupferschmidt & Graham W. Taylor (2023) "Large Language Models" Do Much More than Just Language: Some Bioethical Implications of Multi-Modal AI, The American Journal of Bioethics, 23:10, 110-113, DOI: 10.1080/15265161.2023.2250318

**To link to this article:** https://doi.org/10.1080/15265161.2023.2250318

Published online: 09 Oct 2023.

Visualizations can help democratize discussions of biased AI and further the explainable AI movement: namely, the use of qualitative images, rather than quantitative data, improves communicability and makes epistemic and trust practices possible (Carusi 2008). Two-way public engagement with a diverse group of representatives is key (Bak et al. 2022) especially because images are partially subjective and Transition-Anger will often arise in people who connect most with the (missing) subject, and as such recognize the embedded inequality. Further study is needed on how this is done most effectively, and more generally, on the different ways that AI chatbots may be used to benefit bioethics.

## DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

## FUNDING

The author(s) reported there is no funding associated with the work featured in this article.

## ORCID

Marieke Bak http://orcid.org/0000-0003-0655-0743

## REFERENCES

Bak, M. A. R. 2022. Computing fairness: Ethics of modeling and simulation in public health. *Simulation* 98 (2):103–11. doi:10.1177/0037549720932656.

Bak, M. A. R., V. I. Madai, M. C. Fritzsche, M. T. Mayrhofer, and S. McLennan. 2022. You can't have AI both ways: Balancing health data privacy and access fairly. *Frontiers in Genetics* 13:929453. doi:10.3389/fgene.2022.929453.

Carusi, A. 2008. Scientific visualisations and aesthetic grounds for trust. *Ethics and Information Technology* 10:243–54. doi:10.1007/s10676-008-9159-5.

Cerrato, P., J. Halamka, and M. Pencina. 2022. A proposal for developing a platform that evaluates algorithmic equity and accuracy. *BMJ Health & Care Informatics* 29 (1):1.

Cohen, I. G. 2023. What should ChatGPT mean for bioethics? *The American Journal of Bioethics* 23 (10):8–16. doi:10.1080/15265161.2023.2233357.

Floridi, L. 2010. Ethics after the information revolution. In *The Cambridge handbook of information and computer ethics*, ed. L. Floridi, 3–19. Cambridge: Cambridge University Press.

Lauritzen, P. 2008. Visual bioethics. *The American Journal of Bioethics* 8 (12):50–6. doi:10.1080/15265160802559146.

Mills, C. 2008. Images and emotion in abortion debates. *The American Journal of Bioethics* 8 (12):61–2. doi:10.1080/15265160802559187.

Nussbaum, M. C. 2015. Transitional anger. *Journal of the American Philosophical Association* 1 (1):41–56. doi:10.1017/apa.2014.19.

Ricci Lara, M. A., R. Echeveste, and E. Ferrante. 2022. Addressing fairness in artificial intelligence for medical imaging. *Nature Communications* 13 (1):4581. doi:10.1038/s41467-022-32186-3.

THE AMERICAN JOURNAL OF BIOETHICS
2023, VOL. 23, NO. 10, 110–113
https://doi.org/10.1080/15265161.2023.2250318

Taylor & Francis
Taylor & Francis Group

OPEN PEER COMMENTARIES



# "Large Language Models" Do Much More than Just Language: Some Bioethical Implications of Multi-Modal AI

Joshua August Skorburg[a], Kristina L. Kupferschmidt[a,b], and Graham W. Taylor[a,b]

[a]University of Guelph; [b]Vector Institute for Artificial Intelligence

Cohen (2023) takes a fair and measured approach to the question of what ChatGPT means for bioethics. The hype cycles around AI often obscure the fact that ethicists have developed robust frameworks to address

CONTACT Joshua August Skorburg g.skorburg@gmail.com University of Guelph, Guelph, Canada.

the challenges raised by emerging technologies. That said, the rise of foundation models does represent a paradigm shift whose implications we are only beginning to grasp. Thus, we think Cohen is correct to divide the ethical issues surrounding ChatGPT into those that are "less new to bioethics" and those that are "new-ish."

One shortcoming of Cohen's (2023) approach, however, is the narrow focus on the user-facing, chatbot capabilities of LLMs like ChatGPT. Any ethical analysis centered primarily around the linguistic capabilities of chatbots risks concealing the most promising and worrisome aspect of the underlying technology. Namely, that the transformer architectures powering LLMs have far fewer data-specific "inductive biases" than their predecessors. This means that they can easily consume different kinds of data without much domain-specific engineering, rapidly accelerating their multimodal abilities. With these advances, transformer-based models can fluently translate complex embeddings between different modalities in ways not previously possible. That is, they are not only capable of taking in text as input and generating text as output, but also processing images, audio, code, and video as inputs, and generating these and other modalities as outputs.

Thus, it is unhelpful to treat ChatGPT as the paradigmatic case of cutting-edge AI which may stretch the limits of bioethics. We do not deny that ChatGPT raises some of the issues identified by Cohen (2023). But primarily emphasizing the capabilities of chatbots today is akin to primarily emphasizing the quality of the seats or the placement of the steering wheel in the Ford Model T. In both cases, the narrow focus on the user-facing functionality obscures the bigger-picture, longer-term implications. In automobiles: the shift to (sub)urban lifestyles (among very many others). In bioethics: multimodal agents that extend the potential of LLMs beyond generating well-written prose and programs. Agents are general problem solvers that enable tool use for LLMs. Just like a human using a calculator for math or Google Search for recent information, LLM-based agents use these tools to augment their text-based knowledge.

Before exploring the bioethical implications, a few conceptual clarifications will be helpful. A foundation model is an all-encompassing term to describe modern transformer-based models that use self-supervision and transfer learning at massive scale. A combination of their multi-stage training paradigm based on colossal datasets and the efficient general-purpose transformer architecture has resulted in a previously unexpected level of homogenization, where a single model can be used to complete a multitude of tasks *unanticipated at the time of training*. Importantly, these models are capable of extracting a "universally" generalized embedding that serves as a basis for downstream products and services that can be "unlocked" by further training or even prompting methods such as "in-context learning", which requires no parameter adjustments (Bommasani et al. 2021).

Like all artificial neural networks, transformers require numeric input to extract embeddings. In the case of language, words are converted into lists of numbers called vectors. But this is hardly unique to language. Pixels in an image, voxels in a video, samples in audio, etc. can all be represented as numeric inputs and fed into AI models. For present purposes, the most important property of the transformer architecture is the ability to fluently encode one (or multiple) type/s of input (e.g., text) and decode the embeddings to a different modality output (e.g., images).

What does this mean for medicine? Probably, when most readers think of medical AI, they think about narrow tasks like predicting pneumonia from chest X-rays or melanoma from dermoscopic images. These kinds of models are often trained on a hand-labeled dataset of medical images. Then, when the model sees a new image, it generates a simple positive/negative classification as output. As Moor et al. (2023, p. 259) point out, however, the task is limited by the labels in the training data. These kinds of models often cannot adapt to other closely related tasks, or indeed, even to the same task for a different data distribution, much less write a radiology report.

In contrast, Moor et al. (2023) envision a multimodal Generalist Medical Artificial Intelligence (GMAI). A GMAI is an agent, a general problem solver that would blend inputs from across modalities such as electronic health records, narrative reports, bio signals, medical images, lab results, genomic profiles, and more. In turn, the generated outputs can also be blended across these and other modalities. So rather than a simple positive/negative classification, a GMAI could be queried: "Are those white spots on the image likely to be abscesses?" Or "please highlight the tumor growth since the patient's visit on May 15". Similar to popular tools like DALL-E, text prompts can be used to generate medical images for comparison. Indeed, researchers have raised the possibility of generating protein structures from text prompts. Moor et al. (2023) also imagine a GMAI surgical assistant, capable of responding to queries like, "We cannot find the intestinal rupture. Check whether we missed a view of any intestinal section in the visual feed of the last 15 minutes," and then annotating video streams in real time, carrying out visualization tasks, and even providing spoken information, such as alerts when parts of a procedure are skipped, or providing on-the-spot summaries of relevant medical literature when confronting anatomical rarities during surgery.

This latter example is highly optimistic. If we look outside of the healthcare context though, early examples of tool-using LLMs (e.g., AutoGPT, GPT-Engineer, BabyAGI) are inspiring (if still limited). Moreover, some recent work within the healthcare context does give reason to think that the future of medical AI is likely to be multi-modal. To give just two examples: Soenksen et al. (2022) describe a framework to evaluate medical AI systems that leverage multimodal inputs (tabular data, time-series data, text, and images). The framework consistently produces models that outperform uni-modal approaches across a wide range of clinical domains. Additionally, Ma & Wang (2023) leveraged over 1 million medical images comprising a variety of anatomical structures, pathologies, and medical imaging modalities to produce a universal, medical segment anything model (MedSAM) which outperforms some narrow, specialist medical image segmentation models.

These are just a few examples of what we expect to be a growing research program that harnesses the power of transformers in conjunction with the rich multimodal datasets generated in the medical domain. Building on Cohen's (2023) analysis, then, we suggest that multi-modal AI raises some issues that are "less new" to bioethics, and some that are "new-ish".

Under the "less new" heading, researchers have noted that an important ingredient in the success of LLMs has been access to large, diverse datasets, afforded by the internet (issues of e.g., copyright, fair use, and intellectual property notwithstanding). But it is significantly harder to access large, diverse *medical* datasets. This raises a tension: to reach their full potential, medical foundation models need access to large amounts of sensitive health information, but in a way that does not violate patients' privacy and trust. This makes the classic bioethical issues surrounding consent, privacy, equity, etc. especially salient. Fortunately, frameworks proposed in this journal's pages can provide guidance (McCoy et al., 2023).

Under the heading of "new-ish," we think two issues are especially important.

First, as multimodal approaches become more ubiquitous, the well-known problems of black boxes and bias may become more intractable. An increasingly common practice is to generate text captions from images. These generated captions can then be used to train new models that produce novel image-caption pairs. This autophagous loop is concerning, because if a certain proportion of data used in each generation task is not new, the quality and diversity of outputs degrades. To be clear, the risk for the amplification of biases by machine learning models has been discussed extensively in bioethics. The novelty here is that generated images or other outputs may include "artifacts" from their generator task which will continue to be amplified in each generation loop. Efforts have been made to minimize so-called model collapse, but this often comes at the cost of diversity. One can easily imagine an ML practitioner cherry-picking only the "best" generated outputs to be included in the next generation loop, and this could result in an "echo chamber" effect, where only the most highly represented examples continue through the loops (Alemohammad et al. 2023).

A second issue is what we call the illusion of explanation. Perhaps the most striking feature of chat-enabled LLMs is their ability to converse freely with users. While this is exciting from an accessibility perspective, caution is required. The day is quickly approaching when healthcare professionals may be able to build and use their own standalone models without the help of those with ML-specific training. This raises a "new-ish" problem where users can query models in plain language to, for example, "explain why this treatment plan was chosen," but as Cohen (2023) has rightly pointed out, the responses to such queries may be unreliable hallucinations. Thus, there is an urgent need for new bioethical frameworks to explore the idea of "calibrating" user trust in these model interactions.

There is, of course, much more that could be said. Cohen's analysis is an excellent starting point for these conversations. But addressing the wide range of ethical issues will require bioethicists to look much more broadly than chatbots.

## DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

## FUNDING

The author(s) reported there is no funding associated with the work featured in this article.

## ORCID

Joshua August Skorburg http://orcid.org/0000-0002-3779-5076

Kristina L. Kupferschmidt http://orcid.org/0000-0003-0719-3661
Graham W. Taylor http://orcid.org/0000-0001-5867-3652

## REFERENCES

Alemohammad, S., J. Casco-Rodriguez, L. Luzi, A. Imtiaz Humayun, H. Babaei, D. LeJeune, A. Siahkoohi, and R. G. Baraniuk. 2023. Self-consuming generative models Go MAD. *arXiv preprint arXiv:2307.01850*.

Bommasani, R., D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Cohen, I. G. 2023. What should ChatGPT mean for bioethics? *The American Journal of Bioethics* 23 (10):8–16. doi:10.1080/15265161.2023.2233357.

Ma, J., and B. Wang. 2023. Segment anything in medical images. *arXiv preprint arXiv:2304.12306*.

McCoy, M. S., A. L. Allen, K. Kopp, M. M. Mello, D. J. Patil, P. Ossorio, S. Joffe, and E. J. Emanuel. 2023. Ethical responsibilities for companies that process personal data. *The American Journal of Bioethics* :1–13. doi:10.1080/15265161.2023.2209535.

Moor, M., O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar. 2023. Foundation models for generalist medical artificial intelligence. *Nature* 616 (7956):259–65. doi:10.1038/s41586-023-05881-4.

Soenksen, L. R., Y. Ma, C. Zeng, L. Boussioux, K. Villalobos Carballo, L. Na, H. M. Wiberg, M. L. Li, I. Fuentes, and D. Bertsimas. 2022. Integrated multimodal artificial intelligence framework for healthcare applications. *Npj Digital Medicine* 5 (1):149. doi:10.1038/s41746-022-00689-4.

THE AMERICAN JOURNAL OF BIOETHICS
2023, VOL. 23, NO. 10, 113–115
https://doi.org/10.1080/15265161.2023.2250293



OPEN PEER COMMENTARIES

# Moving from Models to Responsible AI as a Moat

Ashwini Nagappan

University of California, Los Angeles (UCLA)

## INTRODUCTION

In his article, “What Should ChatGPT Mean for Bioethics?” Cohen (2023) highlights novel bioethical issues raised by the emergence of ChatGPT and generative AI more broadly. Among the thought-provoking insights, one stands out to me—the sheer scale of large language models (LLMs) that cements the position of Big Tech companies. While their ownership of these models grants them a seat at the table, it does not guarantee victory in the unfolding generative AI race. Concerns loom over anti-competitive practices, as underscored by the Federal Trade Commission’s (FTC) post emphasizing how control over any of the components underpinning generative AI could influence the competitive landscape (Staff in the Bureau of Competition & Office of Technology 2023). However, understanding who will succeed in the generative AI race transcends anti-competitive practices – the core question is whether generative AI will be deployed responsibly.

## THE MOATS OF TECH BEHEMOTHS AND STARTUPS

Competitive advantages, often referred to as moats, set companies apart from others. The moats wielded by Big Tech incumbents extend beyond their foundational models; they also benefit from access to large datasets, technical expertise, and computing power (Staff in the Bureau of Competition & Office of Technology 2023). Interestingly, a confidential document surfaced on May 4, 2023, indicating that Google believes neither itself nor OpenAI possess any moats (Patel and Ahmad 2023). The document acknowledges the competitiveness of open-source models, yet it downplays incumbents’ ecosystems. Beyond the

CONTACT Ashwini Nagappan ashwininagappan@ucla.edu Department of Health Policy & Management, Fielding School of Public Health, University of California, Los Angeles (UCLA), Los Angeles, CA, USA.