Cambridge Forum on AI: Culture and Society (2026), **2**, e3, 1–10
doi:10.1017/cfc.2025.10013

Cambridge Forum

REFLECTION

A few notes on the scalar foundations of foundation models

Ranjodh Singh Dhaliwal

Faculty of Humanities and Social Sciences, University of Basel, Basel, Switzerland
Email: ranjodh-academia@pm.me; ranjodhsingh.dhaliwal@unibas.ch

(Received 8 March 2025; revised 2 July 2025; accepted 1 December 2025)

Abstract

Foundation models are many things and encompass several modalities; they use text, images, sound, and more recently, action or inference units. But all of these forms share one thing in common: the (massive) scale. The "large" in large language models has been well studied by scholars in critical data, AI and archive studies, with several experts pointing at how these models are environmentally harmful, technically opaque and corporationally monopolistic primarily because of their scale. This piece discusses questions of technical and cultural scale – in the material, archival and procedural senses – within the contemporary technical and discursive landscape. At stake here is the role of critical and design studies within academic, artistic and para-academic worlds. It suggests that instead of corporate chatbots that aspire to pass the Turing test through multipurpose, encyclopedic service, we may be better served by playing with local models and reaching for small-scale AI development. This epistemological shift, in fact, may also provide some creative and critical potential that more effectively gets at the strangeness of machine learning systems while consciously and carefully handling the scalar environmental and social impacts of big AI.

Keywords: critical AI; design; environment and AI; scale

If we were to transpose the famous ancient Asian parable of "The Blind Men and an Elephant" – in which several blind men touch different parts of an elephant and identify it as different animals individually – unto generative AI, then the student will find it to be a homework machine, the technologist might find a hype hope for a future Artificial General Intelligence (AGI)-inflected, Open-AI branded world, your uncle might find a customized fake news generator fit to tell him what he electorally wants to hear, and the economist would find a magical solution to productivity ills he has never been able to (and likely never will) cure. And here, in this corner, we may reach the school of humanistic and social scientific scholars huddled together, yet still touching different parts of the elephant: the optimist will find posthumanist promises, now finally coming to bear, after having been promised ages ago; the educator will look at his/her university memos and sees an academy destroying machine; and the critical scholar will find intellectual property theft, a massive economic extraction from the underpaid global underclass, and a massive infrastructural extraction of water and electricity that contributes to a rapidly warming world (Dhaliwal, 2023; Offert and Dhaliwal, 2024). Now, all of these folks are right in one way or another (except the economist, but that is to be expected, since his whole historical existence is based on being wrong [MacKenzie, 2008; Morgan, 2012; Sweezy, 1970]), but my starting point for this position piece is something that allows this parable to work as one: the fact that the entity at the center is an elephant, an animal of such immense size that it necessarily defies

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (http://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press

the human scale of perception and consequently generates, for the purposes of the parable's narrative structure, these conundrums of subjective knowledge that Hindu, Buddhist and Jain mythologies have found useful over two millennia ago. What happens if we take seriously the scale question at the heart of our differential diagnoses of contemporary artificial intelligence?[1]

Amidst the admirable work done by the nascent field of critical AI studies (Goodlad, 2023; Raley and Rhee, 2023), many scholars have pointed out the extensive environmental toll of this recent scalar expansion – in research, deployment and usage – in generative AI (Anson et al. 2022; Bashir et al. 2024; Bender et al. 2021; Dobbe and Whittaker, 2019; Hao, 2024; Hogan, 2024; Pasek et al. 2023). It is now clear that staggering amounts of electricity and water are needed to power the big data centers that process the petabytes of information scraped from the internet for the calculation of weights that drive these foundation models.[2] These models are designed to be broad spectrum; as several experts studying the economic hype around these models have shown, it would not be unfair to say that, while impressive in parts, these foundation models may be solutions searching for problems (LaGrandeur, 2024; Merchant, 2024; Vinsel, 2021). Casting a wide net here with the sheer scale of foundational models is not only a way for the tech industry to invite use cases and possible directions from all of us (who are now turned into, whether we want or not, the beta testers for such models) but also can be considered as a way of deferring into the future the question of what these models *will* do (instead of can do today). This is also evident from the fact that the business models behind these foundation models have still not proven themselves to be robust and sustainable at scale[3]; OpenAI, the industry leader, for example, despite its massive coffers and billions in revenues, loses more money with each passing week (Gray Widder et al. 2023). It must be noted that this problem of scale – which is publicly only identified by the numerical metrics that denote the number of parameters of a model be it 3 billion or 671, not unlike the traditional question of an archive's scale (Daston, 2017) – is not merely an environmental or business sustainability concern but also an issue regarding ascendant tech monopoly powers (owing to the increasingly absurd costs of training bigger and bigger models [Hosanagar and Krishnan, 2024]), and because these tech monopolies are geographically concentrated, now further a broader global geopolitical (and also data sovereignty and privacy) issue as bigger models start becoming out of reach for most researchers outside the United States of America (Bratton, 2015; Gillespie, 2020; Miller, 2022); consider, for example, the recent trade policy decisions by America which prevent academic/industrial researchers (and even consumers) in most parts of the world from accessing the most advanced integrated circuits which are usually needed to train the state-of-the-art foundation models (Farine, 2025).

If these ubiquitous large foundation models are synonymous, in public discourse, with AI as a whole (Dhaliwal et al. 2024; Schmidt, 2023; Suchman, 2023), then we must consider the very foundations of this equivalence. If we are but beta testers in an unfolding semi-solid global rollout of potentialities, what does that say about our limited agency in the realm of machinic and social collectivities? Unpacking foundation models' scalar dimensions may help answer that question.

To recap, not only are foundational models unimaginably large, but they are also built on a nonconsensual extraction of the labor of everyone in the information economy and are also somehow environmentally catastrophic. If you have been following along any part of critical AI studies, all of this would mark the baseline knowledge of the world today.[4] But the part that I want to focus on

[1]For the purposes of this text, I shall be focusing solely on generative AI, but it is worth noting that AI at large presents problems of scale that must be contended with in a systematic fashion. It must also not be forgotten that the scale issues are not unique to AI and stem from corporate monopolies and big tech (or its big data) in general.

[2]By foundation models, I refer to large, general-purpose models like GPT-4, Claude, Grok and Gemini that are trained on language, images and computational code.

[3]And there is something constitutionally vital about the "model" nature of these two ends of the financial–technical operation (Chadarevian et al. 2004; MacKenzie, 2009).

[4]For more on the political economy of AI within critical AI studies (see Berry, 2025; Burkhardt and Rieder, 2024; Luitse and Denkena, 2021; Narayan, 2022; 2024; Rieder, 2022; Terranova et al. 2024; van der Vlist et al. 2023; van der Vlist, Helmond,

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press

here is the epistemological correlate of that one, one that has to do with the fact that foundation AI models do not have an identity of their own. These are big archival projects – and I am using the term archive here totally in the Derridean, archive studies sense, where an archive anticipates yet unknown applications – that are not merely storing but actively transforming, processing, their contents so that one day some unknown application could make use of the transformation (Jo and Gebru, 2020; Plantin, 2021; Thylstrup, 2022; Thylstrup et al. 2021). And the industry has its fingers crossed really, really hard that the transformation is financially viable. However, we should recall that large models are in fact a small subset of what AI is, has been, and can be, and this piece calls into question our relationships with the scale and size of AI today and tomorrow.

In what follows, I shall here outline a couple of epistemological propositions (or ways of considering) regarding how we may contextualize these scalar conundrums; consider these to be yet more (hopefully interesting) additions to the "what animal is elephant" problem, additions that I hope lead to specific provocations on what is to be intellectually and experimentally done:

1. My first proposition here follows simply the assumption that foundation models can be understood better as objects that invite wonder and thus tie more neatly into our history of fascination than our history of tools and implements. That the foundation models are solutions still searching for a problem to solve is not unrelated with the conceptual framework of techno-solutionism – the idea that social problems can be solved technologically, where the instrumentality of technology is front and center (Kneese, 2023; Richterich, 2024) – here, then, is a fundamental issue plaguing what our models can do. But if the solutionist principle is forcibly excised from the matter, these models come to be seen as material wonders, made of the matter that we are surrounded by, but nevertheless alien in how they appear to us (Amoore et al. 2024; Hayles, 2022). This sense of familiar alienation – and I am here inspired by Katherine Park, Lorraine Daston, Mary Baine Campbell and Caroline Walker Bynum's landmark studies of wonders, miracles and premodern epistemologies (Bynum, 2020; 2015; Campbell, 2011; Daston and Park, 2012) – that the models can invoke in mundane interactions[5] are not only moments of inquiry and investigation but also indicate a moment of pause, a moment of reconsideration of the inter-activity at hand. Treating these objects as material wonders does not mean becoming overwhelmed by their power; even miracles and Christian relics in the late Byzantine era, as Bynum reminds us, were often understood within logical systems of wonder and not always an overpowering theological force. In fact, understanding them as systems that produce surprise in one form or another may help us figure out what surprises us and why.[6] Nothing can be more closely analytical than reading for surprise in mundanity, adopting a methodological stance of what, why and how is our fascination piqued. That is the first step towards cultivating interest that does not take the big, grand sublime as the only wonder possible. If not closely looked at, surprise generation can often be taken as granted to be a feature of "bigness" (Rheinberger, 2011).
2. My second proposition here is that any instrumentality we need/want from our models (however experimental or standard it might be) does not require any foundationality *per se*. In other words, if foundation models are synonymous with large models, be it large language models or large diffusion models, or now multimodal systems of scale, then, I am arguing that anything that we may need the models to "do" (if we need in the first place) does not require systems of

Ferrari, et al. 2024; van der Vlist, Helmond, Luitse, et al. 2024). Several of these studies correctly hint at the "bigness" as a problem, reminiscent of (and drawing from) critical studies of "big" data a decade or so ago.

[5]For example, why did ChatGPT asked several of my students to contact a suicide helpline when they were just asking for the total runtime of Blade Runner? See also, Guzman (2018); Narayanan and Kapoor (2025); Pilling and Coulton (2019) and "Ordinary Media," n.d.

[6]See also, relatedly, special Issue on "Tech | Demo," eds. Julia Eckel, Christoph Ernst and Jens Schröter, *Navi Gationen*, Jg. 23, H. 1 (2023). Relatedly, around the concept of sublime, see Burgess (2023); Castronovo (2019); Mosco (2004).

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press

this magnitude; this is not only true for humanities and social scientific applications[7] but also borne out by emerging work in computer science at large (Chen et al. 2024; Menghani, 2023; Rather et al. 2024; Wu et al., 2024).

A quick look at the history of artificial intelligence shows us that connectionist paradigms – which can be understood as earlier versions of neural network like approaches today – have been conventionally plagued by improper scaling for over half a century now. When artificial neural network (ANN) paradigms were first introduced in the 1960s, their utility and efficiency was severely doubted by the field, and as a result, such paradigms "lost" to the proponents of symbolic AI (think step-by-step predetermined expert systems). And this loss was later recast as a problem of scale; the claim here being that technoscientists did not have enough data or processing power when neural networks were initially proposed. In other words, within the history of technology, it is widely understood that neural networks have always scaled unpredictably (Dhaliwal, 2023; Dhaliwal et al. 2024; Wiggins and Jones, 2023; Jones, 2023; Mendon-Plasek, 2021); most in the 1970s did not predict that powerful machines with a lot of data one day would have ANNs "defeating" the symbolist approaches, just like most experts in 2008 did not realize that something as powerful as ChatGPT could come along only if we scraped the whole internet and tokenized/transformerized it successfully (Daston, 2022; Jones, 2023; Pasquinelli, 2023; Plasek, 2016). These two particular moments[8] are the extrapolatory bedrock of the expansive expansionist approach taken by the tech industry today, the presumption here being that every moment of previous disappointment needed to be followed by scaling more, and as a result, we need to keep on scaling more intensely today, despite the burgeoning evidence to the contrary. It is hardly a universally smart business decision, but since the start of the growth of global productive overcapacity in the 1970s, the economic surplus laying around searching for productive explosions has only taken wi(l)der and wilder speculative bets. Many bets never paid off, but ChatGPT, albeit with major legal IP loopholes underneath it (at least temporarily) did (Benanav, 2019a; 2019b; Koutras and Selvadurai, 2024; Mims, 2023). So, the production of bigger models – precisely what is targeted by the critical scholarship on environment, economy and technology mentioned above – comes from several such parallel bets playing out in real time. Returning to my proposition, I see no reason why *we* must participate in those bets, for we do have much to lose. This is also precisely what extant conceptual (and empirical) critiques of scale and scalability teach us (Hanna and Park, 2020; Lee and Ribes, 2025; Ribes, 2014; tante, 2025; Tsing, 2012; Lempert 2024; Widder and Kim, 2025).

Recent research in computer science has shown signs that larger models may not only be prohibitively costlier but also, in some cases, less likely to sustain their veracity and performance levels in general (Caballero et al. 2022; Kaplan et al. 2020; McKenzie et al. 2023; Zhou et al. 2024). It is imperative then, if we have to (or must) live with AI – and it is a big if, for I am ambivalent about our possibilities – that one thing we *must* do is pick the scale ourselves, carefully, taking full cognizance of our agency as researchers. Small-scale models do very well on many regular tasks; most people do not need the unbounded and imprecise horizon of something as massive as ChatGPT. Smaller, more custom models, especially ones that come pretrained with weights, do not actually present anywhere near the same kind of ecological challenge that the constant expansion done by Microsoft, Oracle, Google, Meta, Anthropic, and OpenAI does; most of the ecological cost of AI comes from training a model and not running it. (And that is not to mention the data privacy cost of having local vs cloud-cartelized models.) So, for the vast variety of tasks, scales much smaller – and consequently, though not always, more sustainable – than the ones wanted by the likes of ChatGPT should/could/would suffice. Plus, given our funding situations and threats globally and the existing archival specificities in our portfolios, rigorous research and custom development of smaller models can actually be done

[7] See, for example, around the context of DH, hints in Berry (2022); Gefen et al. (2021); Hua and Raley (2023); Klein (2022); Bechet et al. (n.d.); Offert and Bell (2020); Smits and Wevers (2023); Berry (2023); Kirschenbaum and Raley (2024).

[8] I can add a third one, perhaps, from ImageNet to image generation in 2012, but establishing that would be beyond the scope of this piece (see Dhaliwal, 2021).

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press

at the cutting edge and unlikely to follow the hype cycles of an extractivist tech industry (which is the only one standing to gain if the models get bigger).

This brings me finally to an interrelated set of design provocations. As much as I have just critiqued the state of solutions searching for a problem, I do so only because of the social relations engendered by ChatGPT as a product, and not because I consider experimentation and failure to be inherently bad, or that the archival deferral is irredeemably problematic (Daston, 2017; Hong, 2020; Strasser and Edwards, 2017). In fact, drawing upon theories of generative failure (Heller, 2008; Marres and McGoey, 2012), in conclusion, I wish to suggest a few possible undercooked directions for humanistic research and exploration with smaller models.

Pardon my polemics and self-cringe but here is how I see the possible scalar queerings of foundational models.[9] On one hand, we may note a scalar engagement actively affecting the epistemic techniques of service and assistance. As several scholars have already noted, service as a cultural technique has been widespread in our ideas about technology; this is why the servers serve us (Canales and Krajewski, 2012; Dhaliwal, 2022; Krajewski, 2018). But not only are there social and agential problems with the idea that our computers must, and exist to, serve us, there is also frankly a lack of interestingness in that notion. The human replacement problem – the idea that AI will or does behave and replace a human, a genealogy older than Turing test (Offert and Dhaliwal, 2024) – is only a problem if "service" is at stake. But if other human modalities, be they creativity or aesthetics or phenomenological experience (Denson, 2023), are to be explored intra-actively (Barad, 2007) then a wider range of what can be done could perhaps open up. On the other hand, scalar shrinkage of these models also perhaps offers us capacities for weirder, queerer, contestations.[10] This includes the ability of – rather, the agency of – different communities and groups and peoples to make their own explorations, with paradoxically both more and less efficient models, or better yet, with efficiencies construed as per the assemblage at hand. So, in other words, I am suggesting models less good at generalized tasks but better at specific or idiosyncratic ones. I don't think our computers should be more like us, but even if that was the case, what could be more like us than bad at most things and good at only one (if that)?

To sum, scale, technically speaking, is often taken by different discourses as a growth framework.[11] Just as conversations around degrowth proliferate out of global material necessity (Saitō, 2022; 2024; Schmelzer et al. 2022; Vettese and Pendergrass, 2022), I want to speculatively suggest that degrowing foundational models might be far more interesting, generative (or even the bare minimum, viable and possible) in our attempting to co/in-habit our human–machinic futures, or to return to the figure I began with, live with an elephant in our midst (Haraway, 2012).

Acknowledgements. I would like to thank the special issue editors (especially Katie MacKinnon, Louis Ravn and Nanna Bonde Thylstrup) and the editorial team for the journal (including Tobias Blanke and Becs Fitchett) for their help and support. For their feedback and insights, my gratitude goes to the anonymous peer-reviewers for the journal and to the hosts, audience and interlocutors (including, but not limited to, Jeffrey Schnapp, Kim Albrecht, Annette Jael Lehmann, Aylin Tschoepe, Tiziana Terranova, Roberta Montinaro, Stamatia Portanova, Teresa Numerico and Salvatore Orlando) at the two venues where previous variants of this work were presented: "Jetsam – Acts of Queering AI" hosted by metalab at the Freie Universität Berlin and "AI and the Quest for Universality: Is Decolonizing AI Possible?" hosted by Centro di Eccellenza Jean Monnet AI-CoDED (Artificial Intelligence and Communication in a Digitalised European Democracy) at the Università degli Studi di Napoli L'Orientale.

Funding statement. N/A

Competing interests. The author declares no known competing interests for this publication.

[9] See, for burgeoning examples, Mark Sample's comments and work on small-scale models (Sample, 2024; "Creative Coding," n.d.). See also, in a different vein, Agre (1998) and Floridi et al. (2021).

[10] See also, relatedly, Förster (2023); Kirschenbaum (2023); Klein (2022); Klein et al. (2025).

[11] For media theory adjacent work on scale as a concept, see DiCaglio (2021); Horton (2021), and relatedly, Morton (2013).

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press

References

Agre, Philip E. 1998. "Toward a critical technical practice: lessons learned in trying to reform AI." In G. Bowker, S. L. Star, L. Gasser & W. Turner (Eds.), *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*, New York: Psychology Press (pp. 131–158).

Amoore, Louise, Alexander Campolo, Benjamin Jacobsen and Ludovico Rella. 2024. "A world model: On the political logics of generative AI." *Political Geography* 113, 103134. https://doi.org/10.1016/j.polgeo.2024.103134

April, Anson, Andrea Ballestero, Dean Chahim et al. 2022. *Water justice and technology: The COVID-19 crisis, computational resource control, and water relief policy*. AI Now Institute at New York University. https://ainowinstitute.org/publication/water-justice-and-technology-report.

Barad, Karen. 2007. *Meeting the universe halfway: quantum physics and the entanglement of matter and meaning*. Duke University Press. https://doi.org/10.1215/9780822388128

Bashir, Noman, Priya Donti, James Cuff, Sroka Sydney, Marija Ilic, Vivienne Sze, Christina Delimitrou and Elsa Olivetti. 2024. "The Climate and Sustainability Implications of Generative AI." *An MIT Exploration of Generative AI*. https://doi.org/10.21428/e4baedd9.9070dfe7.

Bechet, Frederic, Elie Antoine, Jeremy Auguste and Geraldine Damnati. n.d. *Question Generation and Answering for Exploring Digital Humanities Collections*.

Benanav, Aaron. 2019a. "Automation and the Future of Work—1." *New Left Review* 119, 5–38. https://doi.org/10.64590/ifw

Benanav, Aaron. 2019b. "Automation and the Future of Work—2." *New Left Review* 120, 117–146. https://doi.org/10.64590/unr

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA), FAccT'21, 610–623. https://doi.org/10.1145/3442188.3445922.

Berry, David M. 2022. "AI, Ethics, and Digital Humanities." In J. O'Sullivan (Ed.), *The Bloomsbury Handbook to the Digital Humanities* (pp. 125–136). Bloomsbury Handbooks. Bloomsbury Academic.

Berry, David M. 2023. "The Explainability Turn." *Digital Humanities Quarterly* 17, no. 2. https://www.digitalhumanities.org/dhq/vol/17/2/000685/000685.html

Berry, David M. 2025. "Synthetic Media and Computational Capitalism: Towards a Critical Theory of Artificial Intelligence." *AI & SOCIETY* 40, 5257–5269. https://doi.org/10.1007/s00146-025-02265-2.

Bratton, Benjamin H. 2015. *The Stack: On Software and Sovereignty*. Software Studies. MIT Press.

Burgess, Jean. 2023. "Everyday Data Cultures: Beyond Big Critique and the Technological Sublime." *AI & SOCIETY* 38, no. 3: 1243–1244. https://doi.org/10.1007/s00146-022-01503-1

Burkhardt, Sarah and Bernhard Rieder. 2024. "Foundation Models Are Platform Models: Prompting and the Political Economy of AI." *Big Data & Society* 11, no. 2: 20539517241247839. https://doi.org/10.1177/20539517241247839

Bynum, Caroline Walker. 2015. *Christian Materiality: An Essay on Religion in Late Medieval Europe*. (First paperback edition ed.). Zone Books.

Bynum, Caroline Walker. 2020. *Dissimilar Similitudes: Devotional Objects in Late Medieval Europe*. Zone Books.

Caballero, Ethan, Kshitij Gupta, Irina Rish and David Krueger. 2022. "Broken Neural Scaling Laws." Version 17. https://doi.org/10.48550/ARXIV.2210.14891.

Campbell, Mary B. 2011. *Wonder & Science: Imagining Worlds in Early Modern Europe*. Cornell University Press.

Canales, Jimena and Markus Krajewski. 2012. "Little Helpers. About Demons, Angels and Other Servants." *Interdisciplinary Science Reviews* 37, no. 4: 314–331. https://doi.org/10.1179/0308018812Z.00000000025

Castronovo, Russ. 2019. *The Return of the Aesthetic in American Studies*. In W. Fluck J. Völz & R. Jordan (Eds.), REAL - Yearbook of Research in English and American Literature (Vol. 35, pp. 237–260). Narr Francke Attempto Verlag.

Chadarevian, Soraya de, Nick Hopwood and Malcolm Baker. (eds.) 2004. *Models: The Third Dimension of Science*. Original printing. Writing Science. Stanford University Press.

Chen, Dong, Shuo Zhang, Yueting Zhuang, Tang Siliang, Liu Qidong, Wang Hua and Xu Mingliang. 2024. "Improving Large Models with Small Models: Lower Costs and Better Performance." Preprint, arXiv. https://doi.org/10.48550/arXiv.2406.15471.

"Creative Coding." n.d. Accessed July 1, 2025. https://web.sas.upenn.edu/dream-lab/creative-coding-2025/.

Daston, Lorraine ed. 2017. *Science in the Archives: Pasts, Presents, Futures*. The University of Chicago Press.

Daston, Lorraine. 2022. *Rules: A Short History of What We Live By*. The Lawrence Stone Lectures. Princeton University Press.

Daston, Lorraine and Katharine Park. 2012. *Wonders and the Order of Nature: 1150 - 1750*. Zone Books.

Denson, Shane. 2023. *Post-Cinematic Bodies* (Configurations of Film, Vol. 7). Meson press.

Dhaliwal, Ranjodh Singh. 2021. "Rendering the Computer: A Political Diagrammatology of Technology." Ph.D., University of California, Davis. https://www.proquest.com/docview/2562796506/DB239305712D40E1PQ/32.

Dhaliwal, Ranjodh Singh. 2022. "The Cyber-Homunculus: On Race and Labor in Plans for Computation." *Configurations*, 30, no. 4: 377–409. https://doi.org/10.1353/con.2022.0028

Dhaliwal, Ranjodh Singh. 2023. "What Do We Critique When We Critique Technology?" *American Literature* 95, no. 2: 305–319. https://doi.org/10.1215/00029831-10575091

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press

Dhaliwal, Ranjodh Singh, Théo Lepage-Richer and Lucy Suchman. 2024. In *Neural Networks* In Search of Media, Minneapolis; Lüneburg: University of Minnesota Press; meson press.

DiCaglio, Joshua. 2021. *Scale Theory: A Nondisciplinary Inquiry*. University of Minnesota Press.

Dobbe, Roel and Meredith Whittaker. 2019. *AI and Climate Change: How They're Connected, and What We Can Do about It*. Accessed January 3, 2025. https://medium.com/@ainowinstitute.

Farine, Mathilde. 2025. "US Restricts Switzerland's Access to AI Chips." *SWI Swissinfo.Ch*, January 26. Accessed January 3, 2025. https://www.swissinfo.ch/eng/multinational-companies/us-restricts-switzerlands-access-to-ai-chips/88781270.

Floridi, Luciano, Josh Cowls, Thomas C. King and Mariarosaria Taddeo. 2021. "How to Design AI for Social Good: Seven Essential Factors." In *Ethics, Governance, and Policies in Artificial Intelligence*, edited by Luciano Floridi, Springer International Publishing, pp. 125–151. https://doi.org/10.1007/978-3-030-81907-1_9

Förster, Susanne. 2023. "The Bigger the Better?: The Size of Language Models and the Dispute over Alternative Architectures." *A Peer-Reviewed Journal About* 12, no. 1: 120–134. https://doi.org/10.7146/aprja.v12i1.140444

Gefen, Alexandre, Léa Saint-Raymond and Tommaso Venturini 2021. AI for Digital Humanities and Computational Social Sciences. In *Reflections on Artificial Intelligence for Humanity*, edited by Bertrand Braunschweig and Malik Ghallab, (Vol. 12600). Lecture Notes in Computer Science. Springer International Publishing, pp. 191-202. https://doi.org/10.1007/978-3-030-69128-8_12

Gillespie, Tarleton. 2020. "Content Moderation, AI, and the Question of Scale." *Big Data & Society* 7, no. 2: 2053951720943234. https://doi.org/10.1177/2053951720943234

Goodlad, Lauren M. E. 2023. "Editor's Introduction: Humanities in the Loop." *Critical AI* 1, 1–2. https://doi.org/10.1215/2834703X-10734016

Widder, D.G., M. Whittaker and S.M. West. 2023. "Why 'open' AI systems are actually closed, and why this matters." *Nature* 635, 827–833. https://doi.org/10.2139/ssrn.4543807.

Guzman, Andrea L. 2018. "Beyond Extraordinary: Theorizing Artificial Intelligence and the Self in Daily Life." In *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*. Routledge.

Hanna, Alex and Tina M. Park. 2020. "Against Scale: Provocations and Resistances to Scale Thinking." Preprint, arXiv. https://doi.org/10.48550/arXiv.2010.08850.

Hao, Karen. 2024. "Microsoft's Hypocrisy on AI." Technology. *The Atlantic*, September 13. Accessed January 3, 2025. https://www.theatlantic.com/technology/archive/2024/09/microsoft-ai-oil-contracts/679804/.

Haraway, Donna. 2012. *The Companion Species Manifesto: Dogs, People, and Significant Otherness*. 6th print. Paradigm 8 Prickly Paradigm Press.

Hayles, N. Katherine. 2022. "Inside the Mind of an AI: Materiality and the Crisis of Representation." *New Literary History* 54, no. 1: 635–666. https://doi.org/10.1353/nlh.2022.a898324

Heller, Steven. 2008. *Design Disasters: Great Designers, Fabulous Failures, & Lessons Learned*. Allworth Press.

Hogan, Mél. 2024. "The Fumes of AI." *Critical AI* 2, no. 1. https://doi.org/10.1215/2834703X-11205231.

Hong, Sun-ha. 2020. *Technologies of Speculation: The Limits of Knowledge in a Data-Driven Society*. New York University Press.

Horton, Zachary K. 2021. *The Cosmic Zoom: Scale, Knowledge, and Mediation*. The University of Chicago Press.

Hosanagar, Kartik and Ramayya Krishnan. 2024. "Who Profits the Most From Generative AI? - ProQuest." *MIT Sloan Management Review*.

Hua, Minh and Rita Raley. 2023 How to Do Things with Deep Learning Code. In *Digital Humanities Quarterly*. Vol. 17, no. 2. http://www.digitalhumanities.org/dhq/vol/17/2/000684/000684.html.

Jo, Eun Seo and Timnit Gebru. 2020. "Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning." In FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020 New York: Association for Computing Machinery. Barcelona, Spain (pp. 306–316). https://doi.org/10.1145/3351095.33728.

Jones, Matthew L. 2023. "AI in History." *The American Historical Review* 128, no. 3: 1360–1367. https://doi.org/10.1093/ahr/rhad361

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu and Dario Amodei 2020. "Scaling Laws for Neural Language Models." Preprint, arXiv, https://doi.org/10.48550/arXiv.2001.08361.

Kirschenbaum, Matthew. 2023. *Prepare for the Textpocalypse*. March 8. Accessed December 3, 2025. https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318/.

Kirschenbaum, Matthew and Rita Raley. 2024. "AI and the University as a Service." *PMLA/Publications of the Modern Language Association of America* 139, no. 3: 504–515. https://doi.org/10.1632/S003081292400052X

Klein, Lauren. 2022. *Are Large Language Models Our Limit Case?* August 1. https://doi.org/10.5281/ZENODO.6567985.

Klein, Lauren, Meredith Martin, André Brock, Maria Antoniak, Melanie Walsh, Jessica Marie Johnson, Lauren Tilton and David Mimno. 2025. "Provocations from the Humanities for Generative AI Research." Version 1. Preprint, arXiv. https://doi.org/10.48550/ARXIV.2502.19190.

Kneese, Tamara. 2023. *Death Glitch: How Techno-Solutionism Fails Us in This Life and Beyond*. Yale University Press.

Koutras, Nikos and Niloufer Selvadurai eds. 2024. *Recreating Creativity, Reinventing Inventiveness: Challenges Facing Intellectual Property Law*. Law in Times of Crisis. Routledge, Taylor and Francis Group.

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press

Krajewski, Markus. 2018. *The Server: A Media History from the Present to the Baroque*. Translated by Ilinca Iurascu, Yale University Press.

LaGrandeur, Kevin. 2024. "The Consequences of AI Hype." *AI and Ethics* 4, no. 3: 653–656. https://doi.org/10.1007/s43681-023-00352-y

Lee, Francis and David Ribes. 2025. "Computational Universalism, or, Attending to Relationalities at Scale." *Social Studies of Science*, July 14, 03063127251345089. https://doi.org/10.1177/03063127251345089

Lempert, Michael. 2024. *From Small Talk to Microaggression: A History of Scale*. University of Chicago Press.

Luitse, Dieuwertje and Wiebke Denkena. 2021. "The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI." *Big Data & Society* 8, no. 2: 20539517211047734. https://doi.org/10.1177/20539517211047734

MacKenzie, Donald A. 2008. *An Engine, Not a Camera: How Financial Models Shape Markets*. First MPI Press. paperback. Inside Technology. MIT Press.

MacKenzie, Donald A. 2009. In Material Markets: How Economic Agents Are Constructed Clarendon Lectures in Management Studies. Oxford University Press.

Marres, Noortje and Linsey McGoey. 2012. "Experimental Failure: Notes on the Limits of the Performativity of Markets." Paper presented at After Markets: Researching Hybrid Arrangements. https://research.gold.ac.uk/id/eprint/7353/.

McKenzie, Ian R., Alexander Lyzhov, Michael Pieler, Alicia Parrish, Aaron Mueller, Ameya Prabhu, Euan McLean, Aaron Kirtland, Alexis Ross, Alisa Liu, Andrew Gritsevskiy, Daniel Wurgaft, Derik Kauffman, Gabriel Recchia, Jiacheng Liu, Joe Cavanagh, Max Weiss, Sicong Huang, The Floating Droid, Tom Tseng, Tomasz Korbak, Xudong Shen, Yuhui Zhang, Zhengping Zhou, Najoung Kim, Samuel R.Bowman and Ethan Perez. 2023. "Inverse Scaling: When Bigger Isn't Better." *Transactions on Machine Learning Research*, June 25. https://openreview.net/forum?id=DwgRm72GQF.

Mendon-Plasek, Aaron. 2021. "Mechanized Significance and Machine Learning: Why It Became Thinkable and Preferable to Teach Machines to Judge the World." In *The Cultural Life of Machine Learning*, edited byJonathan Roberge and Michael Castelle, Springer International Publishing, pp. 31–78. https://doi.org/10.1007/978-3-030-56286-1_2

Menghani, Gaurav. 2023. "Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better." *ACM Computing Surveys* 55, no. 12: 259:1–259:37. https://doi.org/10.1145/3578938

Merchant, Brian. (2024). *AI Generated Business: The Rise of AGI and the Rush to Find a Working Revenue Model*. AI Now Institute. https://ainowinstitute.org/general/ai-generated-business

Miller, Chris. 2022. *Chip War: The Fight for the World's Most Critical Technology*. First Scribner hardcover edition. Scribner, an imprint of Simon & Schuster.

Mims, Christopher. 2023. "AI Tech Enables Industrial-Scale Intellectual-Property Theft, Say Critics." Tech. *Wall Street Journal*, February 4. Accessed January 5, 2025. https://www.wsj.com/articles/ai-chatgpt-dall-e-microsoft-rutkowski-github-artificial-intelligence-11675466857.

Morgan, Mary S. 2012. *The World in the Model: How Economists Work and Think*. Cambridge University Press. https://doi.org/10.1017/CBO9781139026185

Morton, Timothy. 2013. *Hyperobjects: Philosophy and Ecology after the End of the World*. Posthumanities 27 University of Minnesota press.

Mosco, Vincent. 2004. *The Digital Sublime: Myth, Power, and Cyberspace*. MIT press.

Narayan, Devika. 2022. "Platform Capitalism and Cloud Infrastructure: Theorizing a Hyper-Scalable Computing Regime." *Environment and Planning A: Economy and Space* 54, no. 5: 911–929. https://doi.org/10.1177/0308518X221094028

Narayan, Devika. 2024. "Cloud as Infrastructure: Theorising the Links Between 'Big' Tech and 'Small' Tech." *AoIR Selected Papers of Internet Research* ahead of print. https://doi.org/10.5210/spir.v2024i0.14012.

Narayanan, Arvind and Sayash Kapoor. 2025. "AI as Normal Technology." *Knight First Amendment Institute* 25, no. 09. https://perma.cc/HVN8-QGQY

Offert, Fabian and Peter Bell. 2020. "Generative Digital Humanities." Paper presented at CHR 2020: Workshop on Computational Humanities Research, Amsterdam. *CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020*, *Amsterdam*, *The Netherlands*. https://ceur-ws.org/Vol-2723/short23.pdf.

Offert, Fabian and Ranjodh Singh Dhaliwal. 2024. "The Method of Critical AI Studies, A Propaedeutic." Version 1. Preprint, arXiv. https://doi.org/10.48550/ARXIV.2411.18833.

"Ordinary Media |." n.d. Accessed July 2, 2025. https://sites.northwestern.edu/ordinary/.

Pasek, Anne, Cindy Kaiying Lin, Zane Griffin Talley Cooper and Jordan B. Kinder. 2023. "Digital Energetics." In *Search of Media*, Meson Press : University of Minnesota Press.

Pasquinelli, Matteo. 2023. *The Eye of the Master: A Social History of Artificial Intelligence*. Verso.

Pilling, Franziska and Paul Coulton. 2019. "Forget the Singularity, Its Mundane Artificial Intelligence That Should Be Our Immediate Concern." *The Design Journal* 22, no. sup1: 1135–1146. https://doi.org/10.1080/14606925.2019.1594979

Plantin, Jean-Christophe. 2021. "The Data Archive as Factory: Alienation and Resistance of Data Processors." *Big Data & Society* 8, no. 1: 20539517211007510. https://doi.org/10.1177/20539517211007510

Plasek, Aaron. 2016. "On the Cruelty of Really Writing a History of Machine Learning." *IEEE Annals of the History of Computing* 38, no. 4: 6–8. https://doi.org/10.1109/MAHC.2016.43

Raley, Rita and Jennifer Rhee. 2023. "Critical AI: A Field in Formation." *American Literature*, 95, no. 2: 185–204. https://doi.org/10.1215/00029831-10575021

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press

Rather, Ishfaq Hussain, Sushil Kumar and Amir H. Gandomi. 2024. "Breaking the Data Barrier: A Review of Deep Learning Techniques for Democratizing AI with Small Datasets." *Artificial Intelligence Review* 57, no. 9: 226. https://doi.org/10.1007/s10462-024-10859-3

Rheinberger, Hans-Jörg. 2011. "Infra-Experimentality: From Traces to Data, from Data to Patterning Facts." *History of Science* 49, no. 3: 337–348. https://doi.org/10.1177/007327531104900306

Ribes, David. 2014. "Ethnography of Scaling, or, How to a Fit a National Research Infrastructure in the Room." *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (New York, NY, USA), CSCW'14, February 15, 158–170. https://doi.org/10.1145/2531602.2531624.

Richterich, Annika. 2024. "Can't Fix This? Innovation, Social Change, and Solutionism in Design Thinking." *Media and Communication* 12. https://doi.org/10.17645/mac.7427

Rieder, Bernhard. 2022. "Towards a Political Economy of Technical Systems: The Case of Google." *Big Data & Society* 9, no. 2: 20539517221135162. https://doi.org/10.1177/20539517221135162

Saitō, Kōhei. 2022. *Marx in the Anthropocene: Towards the Idea of Degrowth Communism*. Cambridge University Press.

Saitō, Kōhei. 2024. *Slow down: The Degrowth Manifesto* (First edition). Astra House.

Sample, Mark. 2024. "I really need to publish a keynote I delivered in the spring on the pleasures of hand-crafting 'small language models'—and I specifically…." *Post. Bluesky* December 30. https://bsky.app/profile/samplereality.bsky.social/post/3lejwlspyv22j.

Schmelzer, Matthias, Andrea Vetter and Aaron Vansintjan. 2022. *The Future Is Degrowth: A Guide to a World beyond Capitalism*. Verso.

Schmidt, Benjamin. 2023. "Representation Learning." *The American Historical Review* 128, no. 3: 1350–1353. https://doi.org/10.1093/ahr/rhad363

Smits, Thomas and Melvin Wevers. 2023. "A Multimodal Turn in Digital Humanities. Using Contrastive Machine Learning Models to Explore, Enrich, and Analyze Digital Visual Historical Collections." 38, no. 3: 1267–1280. https://doi.org/10.1093/LLC/FQAD008

Strasser, Bruno J. and Paul N. Edwards. 2017. "Big Data Is the Answer … But What Is the Question?" *Osiris* 32, no. 1: 328–345. https://doi.org/10.1086/694223

Suchman, Lucy. 2023. "The Uncontroversial 'Thingness' of AI." *Big Data & Society* 10, no. 2: 20539517231206794. https://doi.org/10.1177/20539517231206794

Sweezy, Paul M. 1970. "Toward a Critique of Economics." *Review of Radical Political Economics* 2, no. 1: 1–8. https://doi.org/10.1177/048661347000200101

tante. 2025. "Against Scale." *Smashing Frames*, January 26. Accessed February 7, 2025. https://tante.cc/2025/01/26/against-scale/.

Terranova, Tiziana and Daniël de Zeeuw, and **Institute of Network Cultures**. 2024. *Undead Digital Labor and the General Intellect – A Conversation on AI between Tiziana Terranova and Daniël de Zeeuw*. April 23. Accessed February 7, 2025. https://networkcultures.org/blog/2024/04/23/terranova-and-de-zeeuw/.

Thylstrup, Nanna Bonde. 2022. "The Ethics and Politics of Data Sets in the Age of Machine Learning: Deleting Traces and Encountering Remains." *Media, Culture & Society* 44, no. 4: 655–671. https://doi.org/10.1177/01634437211060226

Thylstrup, Nanna Bonde, Daniela Agostinho, Annie Ring, Catherine D'Ignazio and Kristin Veel eds. 2021. *Uncertain Archives: Critical Keywords for Big Data*. MIT Press.

Tsing, Anna Lowenhaupt. 2012. "On Nonscalability: The Living World Is Not Amenable to Precision-Nested Scales." *Common Knowledge* 18, no. 3: 505–524. https://doi.org/10.1215/0961754X-1630424

van der Vlist, Fernando N, Anne Helmond and Fabian Ferrari. 2023. "Big AI: The Cloud as Marketplace and Infrastructure." *AoIR Selected Papers of Internet Research* ahead of print. December 31 https://doi.org/10.5210/spir.v2023i0.13510.

van der Vlist, Fernando N, Anne Helmond and Fabian Ferrari. 2024. "Big AI: Cloud Infrastructure Dependence and the Industrialisation of Artificial Intelligence." *Big Data & Society*, 11, no. 1: 20539517241232630. https://doi.org/10.1177/20539517241232630

van der Vlist, Fernando N, Anne Helmond, Dieuwertje M. R. Luitse, Bernhard Rieder, Sam Hind and Max Kanderske. 2024. "The Political Economy of AI as Platform: Infrastructures, Power, And The AI Industry." *AoIR Selected Papers of Internet Research* ahead of print. https://doi.org/10.5210/spir.v2024i0.14088.

Vettese, Troy and Drew Pendergrass. 2022. *Half-Earth Socialism: A Plan to Save the Future from Extinction, Climate Change, and Pandemics*. Verso.

Vinsel, Lee. 2021. "You're Doing It Wrong: Notes on Criticism and Technology Hype." Medium, February 1. Accessed January 3, 2025. https://sts-news.medium.com/youre-doing-it-wrong-notes-on-criticism-and-technology-hype-18b08b4307e5.

Widder, David Gray and Nathan Kim. 2025. "How Big Cloud Becomes Bigger: Scrutinizing Google, Microsoft, and Amazon's Investments." SSRN Scholarly Paper No. 5377426. Social Science Research Network, July 31. https://doi.org/10.2139/ssrn.5377426.

Wiggins, Chris and Matthew L. Jones. 2023. How Data Happened: A History from the Age of *Reason to the Age of Algorithms* First edition. W. W. Norton & Company.

Wu, Carole-Jean, Bilge Acun, Ramya Raghavendra and Kim Hazelwood. 2024. "Beyond Efficiency: Scaling AI Sustainably." *IEEE Micro* 44, no. 5: 37–46. https://doi.org/10.1109/MM.2024.3409275

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press

Zhou, Lexin, Wout Schellaert, Fernando Martínez-Plumed, Yael Moros-Daval, Cèsar Ferri and José Hernández-Orallo. 2024. "Larger and More Instructable Language Models Become Less Reliable". *Nature* 634, no. 8032: 61–68. https://doi.org/10.1038/s41586-024-07930-y

Ranjodh Singh Dhaliwal is a professor of Digital Humanities, Artificial Intelligence and Media Studies in the department of Arts, Media, and Philosophy at the University of Basel, where he also directs the Digital Humanities Laboratory. He was previously the Ruth and Paul Idzik Collegiate Chair in Digital Scholarship, English, and Film, Television, and Theater at the University of Notre Dame. He is the co-author (with Théo Lepage-Richer and Lucy Suchman) of *Neural Networks* (University of Minnesota Press and meson press, 2024), and his award-winning writing – situated between media theory, literary studies, computer science, political economy, critical design and STS – can be found in *Critical Inquiry, Configurations, Social Text, American Literature, Journal of Cinema and Media Studies, ACM FDG, ACM UIST, Design Issues* and *Frankfurter Allgemeine Zeitung*, among other scholarly and popular venues. He is an incoming president of the Society for Literature, Science, and the Arts (SLSA).

Cite this article: Dhaliwal, R. S. 2026. A few notes on the scalar foundations of foundation models. *Cambridge Forum on AI: Culture and Society*, **2**, e3, 1–10. https://doi.org/10.1017/cfc.2025.10013

https://doi.org/10.1017/cfc.2025.10013 Published online by Cambridge University Press