Jun 8th, 9:00 AM - Jun 12th, 5:00 PM

Small AI: A degrowth imaginary for designing with/for artificial intelligence

Roy Bendor
Delft University of Technology, The Netherlands

Dave Murray-Rust
Delft University of Technology, The Netherlands

Rainer Rehak
Weizenbaum Institute, Germany

Citation

Bendor, R., Murray-Rust, D., and Rehak, R. (2026) Small AI: A degrowth imaginary for designing with/for artificial intelligence, in Simeone, L., Gray, C. M., Verhoeven, A., de Götzen, A., Bakırlıoğlu, Y., Zohar, H., Stead, M., and Buwert, P. (eds.), *DRS2026: Edinburgh*, 8–12 June, Edinburgh, United Kingdom.
https://doi.org/10.21606/drs.2026.1193

Small AI: A degrowth imaginary for designing with/for artificial intelligence

Roy Bendor[a], Dave Murray-Rust[a], Rainer Rehak[b]

[a] Faculty of Industrial Design Engineering, Delft University of Technology, The Netherlands
[B] Research group Digitalization, Sustainability, and Participation, Weizenbaum Institute, Germany

*Corresponding author e-mail: r.bendor@tudelft.nl

doi.org/10.21606/drs.2026.1193

Abstract: This paper presents Small AI as a conceptual lever for designers to resist the 'bigger is better' trend in the discourse and development of Generative AI. We explore different types of Small AI: initially in relation to the size of models, as smaller models require less computational resources for both training and inference, thus implying a smaller ecological footprint. However, we extend the meaning of smallness to apply to political, epistemic and cultural domains as well. In this vein, the paper posits Small AI as an imaginary-in-the-making, and a rallying concept for an alternative innovation agenda that holds, in the spirit of E.F. Schumacher's famous dictum, that it is small (AI) and not big (AI) that is beautiful. The paper draws conclusions for those designing with/for AI, positing Small AI as a form of 'unmaking' with which designers may re-articulate the entangled relations between users, computation, and worlds.

Keywords: sustainability; critical design; conviviality; generative AI

1. Introduction

The trajectory of artificial intelligence (AI) as a sociotechnical phenomenon can be seen as typical of late-stage capitalism. The means of algorithmic computation and the copious amounts of (human-created) data absorbed by AI systems are supported by, and prop up, environmental extractivism, labour exploitation and deskilling, and all-encompassing cultural appropriation and commodification (McQuillan, 2022; Pasquinelli, 2023; Zuboff, 2019). The creation of AI systems appears vampiric, a solution looking for a problem while depleting the resources of a world it claims to save. As currently practiced, the growth of the AI industry spells not only a crisis for computation but one for society and the planet. Through this lens, computation appears as a form of defuturing (Fry, 2020), epistemicide (Couldry & Mejias, 2019), or ecocide (Comber & Eriksson, 2023).

Big Tech's approach to developing AI systems relies on massive investments in computational infrastructure – a veritable computational arms race floated by a manufactured sense of necessity and inevitability (Bareis & Katzenbach, 2022). This "Gold Rush" (Virz & Huang, Feb.2, 2025) mentality is evident not only in the scope of financial investments the industry has garnered from both private sector and government – estimates range from USD$110B to USD$136B in 2024 alone (Lifset et al., 2025; Lunden, Feb.11, 2025) – but also in the kind of metrics it uses for measuring the performance of large language models (Bender et al., 2021). The insatiable drive to scale globally by developing faster processors (implicated in international trade wars and neocolonial practices) and building behemoth data centers (possibly energized by bespoke nuclear reactors) is mere material evidence for a technology whose hallmark is 'bigger is better' – no matter the costs (Varoquaux, Luccioni & Whittaker, 2025). The social and environmental implications of Extra-Large AI are already visible: centralized, profit-driven development of ever-larger models, vague business models rich with magical promises (Rehak, 2025), and with an enormous social and environmental toll (Bender et al., 2021; Brevini, 2022; Crawford, 2021; Luccioni, Jernite & Strubell, 2024; Schütze, 2024).

Responses to the direction taken by the AI industry range from acquiescence – 'it's inevitable so why bother? – to outright rejection – 'until society changes, AI will remain evil'. Since neither response is tenable (AI is not going away anytime soon, and the current direction taken by the industry is abhorrent to say the least), this paper is motivated by our desire to to start charting an alternative path – one that goes beyond reform to offer a transformational perspective on AI. The challenge, as we see it, involves developing new imaginaries that frame practices that may bring alternative paths to life (Bendor, 2025). AI has always been a metaphor-driven field (Agre, 1997; Rehak, 2025), where the terms used to describe AI systems have the power to mobilize and align different actors (Blythe, Lindley & Murray-Rust, 2025; Dove & Fayard, 2020; Murray-Rust, Nicenboim & Lockton, 2022). Whether we cast AI systems as industrial components (Lindley et al., 2020), various forms of magical production and sleight of hand (Derczynski, 2023; Lupetti & Murray-Rust, 2024), a fuzzy synoptic view of culture (Chiang, 2023), or provoke images of its semantic void (Bender et al., 2021), the terms we use to think with, the way we do metaphoring (Nolan, 2021) and construct imaginaries set up the possibilities for considered action.

In this conceptual paper we trace the emergence of a new imaginary-in-the-making for Generative AI (GenAI in short). We refer to it broadly as 'Small AI' because its defining characteristic is the much smaller size of the language models (LMs) it utilizes. These require less computational demands and are therefore less energy intensive to work with. But as we argue below, smallness can be extended to apply to various political, epistemic and cultural domains as well, and in this sense may align better with the principles of degrowth for "The real conflict is not between technology and anti-technology. It is about how technology is imagined and the conditions under which it is deployed" (Hickel, 2023). If there is a role for digital technology in general (Kwet, 2024), and GenAI in particular, to play in the advancement of a degrowth agenda, the question is, in which form (if at all) can its development and use be justified from *both* social and environmental perspectives?

Small AI offers a starting point to answer this question. Small language models are not only less resource hungry but allow for a much more distributed type of computation, potentially eroding the power of current large players to dominate the field, dictate regulatory regimes, set standards, and control the flow of information in and through AI models (Ahmed, Wahed & Thompson, 2023; Allen & Weyl, 2024; Widder, Whittaker & West, 2024). As such, Small AI may spell the end of the current AI development paradigm and usher in an era of "algorithmic pluralism" (Jain et al., 2024), offering a new sociotechnical imaginary for GenAI, complete with a convivial design agenda that flips the current model on its head and suggests, much in the spirit of E.F. Schumacher's (1973) famous dictum, that it is small (AI) and not big (AI) that is beautiful (Rehak, 2026).

2. Small AI

Using smallness as a defining characteristic of GenAI systems presents a challenge. On the one hand, it invites an investigation into a range of different types of smallness that exceed any specific trait and reveal GenAI as a complex, entangled socio-enviro-political phenomenon (Bridle, 2022; Frauenberger, 2019; Rakova & Dobbe, 2023). On the other hand, ascertaining what exactly constitutes smallness when it comes to an LM is neither purely technical, straightforward nor precise, but rather an exercise in both material and moral argumentation. In what follows we start by describing some of the more familiar types of smallness, and gradually expand the repertoire. We hope the reader will treat the overlaps between some forms of smallness not as a weakness, that is, as evidence of the framework's lack of analytical clarity or specificity, but as an invitation to dig further into the entanglements that characterize AI, and which render it both exciting and daunting.

What makes our contribution potentially more impactful than other, kindred concepts such as Green AI, Frugal AI, Slow AI, or Responsible AI – to name but a few – is that *we assert an inherent, fundamental connection between the ecological implications and social relevance of GenAI applications*. Green AI (Schwartz et al., 2020) focuses on the relations between accuracy and efficiency and offers tools to reduce AI's ecological footprint. Similarly, Frugal AI (Girdhar et al., 2025) takes aim at the ecological implications of AI but extends the scope to address issues around affordability and scalability. Slow AI (Crampton, 2020) offers heuristics for AI use (and nonuse) from a strictly political perspective that emphasizes critical reflection and resistance. And Responsible AI (Dignum, 2019) attempts to chart a path for AI use that complies with moral and ethical imperatives. Small AI, in contrast, offers researchers, designers and technologists an integrated approach that can serve as a rallying term and lens for a variety of perspectives that may ultimately consolidate alternative sociotechnical practices and imaginaries.

2.1 Smallness as a measure of ecological footprint

In our formulation, smallness may mean several different things, but this polysemy builds on one, *sine qua non* characteristic: Small AI relies on technical processes and infrastructures that require less resources and are more locally rooted. Ecological smallness can be measured in terms of joules, CO2 equivalents, litres of water, tonnes of rare earth minerals, and so on, but this is not straightforward. Evaluating the full environmental lifecycle of any technology is challenging (Istrate et al., 2024; Kaack et al., 2022; Rohde et al., 2024), even more so when producers and vendors – with very few exceptions (eg. Mistral AI, July 22, 2025) – keep the data confidential (Luccioni, Jernite & Strubell, 2024) or provide suspect data (O'Brien, Sept.15, 2024). So, while recognition of the growing ecological footprint of AI has become widely acknowledged (De Vries, 2023), the success of efforts to reduce said footprint have been rather limited (Nenno, 2024).

The ecological footprint of GenAI comprises both hardware and software demands, and may be informed by the number of model parameters or the size of the data used to train models (aspects we return to below), as well as the efficiency of the model. It should consider the cost of the training process, the ratio between training and inference, the amount of inferences made in typical usage, and more (Bai et al., 2024; Luccioni, Jernite & Strubell, 2024). For example, while a typical inference (prompt) directed to ChatGPT requires an estimated 20 Wh (Smith, Oct.1, 2025), small models use orders of magnitude less. This sense of ’typical usage’ is one of the key reasons for using ’small’ rather than, say ’efficient’ or another technically grounded term: it remains open to questions of both imaginaries and usage.

2.2 Smallness as a measure of model parameters

The smallness of an LM may be derived from the number of model parameters, understood as those controls by which models ‘learn’ about the world – how different data are statistically connected, and how these connections are assigned with significance, weight, and so forth. On the surface, the more parameters in a GenAI model the better it can address complex relationships in the world. ChatGPT5, for instance, is estimated to include between several to tens of trillions of parameters (Comet API, Oct.18, 2025). But the more parameters the higher the cost of training and inference (in both time and energy) and the more non-contextualised the connections are (De Vries, 2023; Luccioni, Jernite & Strubell, 2024; Patterson et al., 2021). It follows that models with fewer parameters would typically, even if not necessarily, result in a smaller ecological footprint and, possibly, better performance in a narrow, specific domain. For example, a model that includes 7 billion parameters (such as LLaMa-7B), can be run on a high spec 3 year old laptop, which opens up many possibilities for local use. Applications using TinyML or EdgeML ‘distil’ large models into quantized, reduced forms that are small enough to run on microprocessors, IoT devices and standard laptops (Lin et al., 2023). Miniaturization does not necessarily mean inferior performance. Language models with fewer parameters may outperform larger ones if trained “deeply” (Hoffmann et al., 2022).

This kind of smallness is often desirable for its own sake: if a model fits on a laptop graphics card, it can be used locally, alleviating worries about sharing data with model providers. If it runs on a lightweight microprocessor, it can be built into IoT devices, used in small spaces, powered by batteries and so on. Smallness here is an enabling factor, an opening of possibilities, as well as reduction of harms. At the same time, the ability to run local LMs in a relatively cheap and energy efficient manner raises the prospect of rebound effects (Fouquet & Hippe, 2022), effectively trading gains in material reduction for increased use.

2.3 Smallness as a measure of data intake

Another way to shrink the size (and, consequently, the ecological footprint) of models has to do with the amount of data they use. Current LMs often utilize the “Common Crawl” dataset to absorb a significant proportion of everything ever put on the internet. This results in large datasets – the latest crawl in Oct. 2025 yielded roughly 514 terabytes of uncompressed content (Çelikkanat, Oct.25, 2025) – that not only ignore common copyright protections and are difficult to clean up, standardize and manage, but also exceed the data collection and analysing capacity of any individual or small group, thus reinforcing the centripetal tendencies of the technology (we return to this below). For example, while the size of datasets used to train OpenAI's ChatGPT is kept secret, it is estimated that GPT3 was trained with about 45 terabytes of data (Brown et al., 2020).

One approach to data smallness includes careful curation and pruning. A RAVE audio model (Caillon & Esling, 2021) that produces sound in a particular flavour can be trained with only 1-10 hours of audio, a reasonable corpus for a musical practitioner to generate on their own, and a model size that can easily fit on a personal computer instead of running on a cloud server. Similarly, 'indigenous AI' – a distributed set of practices that aim to preserve languages with less than a thousand speakers, hence with minimal training corpora – makes use of carefully curated data that offer a better outcome than a maximalist approach (Pinhanez et al., 2024). Another approach to data smallness is to reuse (parts of) large models in novel ways. In such cases the ability to work with small datasets for fine tuning, or even 'zero shot' abilities to work without additional training data at all, depends on previous training on larger datasets. The sharing of open models by projects such as HuggingFace and Ollama mean that the upfront 'largeness' of training a complete language model makes available the possibility of smallness for others.

2.4 Smallness as a measure of distributedness

The centripetal tendencies of the AI industry have been widely acknowledged, raising concerns about the growing ability of the large players in the GenAI field to consolidate and leverage monopolistic practices (Jain et al., 2024; Khanal, Zhang & Taeihagh, 2025; Wach et al., 2023; Widder, Whittaker & West, 2024). Echoing earlier critiques of the mass media (Baker, 2006) and social media (Fuchs, 2015), such practices tend to undermine government oversight by heavily influencing relevant policymaking. But they also increase the temptation to apply invasive surveillance, and curate and deploy models that promote the interests of Big Tech by distributing disinformation and obfuscating the ways in which this takes place.

Small AI represents possibilities to undermine this tendency by providing individuals and communities with a measure of freedom to experiment with GenAI while maintaining privacy, agency and autonomy. Much in the spirit of Illich's (1973) notion of conviviality, civil society actors can go beyond merely accessing large models in decentralized ways to actually train and run GenAI models by themselves and for themselves. GenAI, in this way, can become a means for community creativity and self-actualization through computation (Engelbutzeder et al., 2025), instead of serving the alienating forces of late-stage capitalism (Steig et al., 2025). If we imagine the range of a model as consisting of all the tokens it holds – who and what it connects, and in which ways – the smaller a model's range, the more distributed its operation is and the more opportunities it offers for local instances, specific customizations, opportunities for convivial uses, and possibilities to pluralize human-algorithm interactions (Jain et al., 2024).

Furthermore, by making models truly open source, sharing computing resources (as with Gensyn), or hacking corporate models (as with LLaMA derivatives such as Vicuna and WizardLM), communities are finding innovative ways to reject the power of Big Tech to monopolize the creation and use of GenAI. Whereas 'openness' when it comes to large language models (LLMs) is often a misnomer (Widder, Whittaker & West, 2024), applications made 'in the small' enable more direct and meaningful contestation in terms of proximity to the model and its uses (Rehak, Ullrich & Vladova, 2025). The range of projects initiated by the AI x Design community (https://aixdesign.co/) and the distributed manner by which the community manages itself are testament to that.

2.5 Smallness as a measure of pluriversality

One of the hallmarks of all GenAI applications is their ability to move between instances and generalized patterns. Generalization, however, may quickly turn into forced universality. Ask, for example, Meta's AI who is the President, and the answer will default to the president of the USA. In a perverse version of the one-size-fits-all model, what is relevant for Big Tech is taken to be relevant for everyone. Yet, it is possible to take a more situated perspective into model inputs and outputs, suggesting ways to ground models in a particular culture, community, site of production, and so forth, and in this way gesture away from totalizing claims to model *the* world (Amoore et al., 2024) onto humbler aims to model *a* world – one amongst many. In this sense, while cultural smallness may sound initially like a reductive solution, it can also be seen as a marker of pluriversality (Escobar, 2018), one that may safeguard difference, promote a respect for particularity and offer a deviation from the homogenizing, global average of blandness often peddled by LLMs (Sourati, Ziabari & Dehghani, 2026).

Machado de Oliveira's experimentation with ChatGPT hints at the possibility that even a run-of-the-mill GenAI tool may promote pluriversality and meta-relationality (Machado de Oliveira & Cinnamon Tea, 2025), although, in this case, the use of reinforcement learning leaves the largeness of the model intact and thus falls short of qualifying as an instance of Small AI. We can, however, find smaller LMs that reject universalism to better serve the needs of a particular culture. For instance, GPT-NL (https://gpt-nl.nl/) is trained exclusively on documents in Dutch and is committed not only to ethical data sourcing but also to reducing the model's environmental footprint. In a similar spirit, the multilinguistic platform Papa Rao (https://papareo.nz/) aims to protect indigenous linguistic cultures (primarily Māori) based on "guardianship rather than ownership of data".

3. Small AI as an invitation to unmake designing with/for AI

In the previous section we have articulated the main characteristics of Small AI, pointing to alternative practices that challenge the prevailing 'bigger is better' imaginary. But what could or should designers do to promote Small AI? In this section we describe the underlying task of designers as one of "unmaking" (Jones, 2021), that is, an opportunity for designers to reflect on their relations to GenAI and, in response, reconfigure the web of relations (and meanings) that GenAI holds together.

It can be argued that simply introducing new practices will not necessarily displace capitalist structures as long as the imaginaries upon which the industry is built remain intact (Feola, 2019). Reconfiguring *both* practice and imaginaries can be seen as a form of unmaking – the redoing of "harmful relationships that have emerged in the aftermath of previous makings" (Lindström & Ståhl, 2020, p. 12). Unmaking can be an emancipatory, agonistic move (Sabie et al., 2022, 2023; Lupetti & Murray-Rust, 2024) that helps to redraw relations and provide spaces, niches and cracks for new and challenging practices and relationalities to emerge. Lindström & Ståhl (2023) explain this as a bi-directional movement where one relation is unmade to give space for another. The plurality of forms of smallness we have illustrated above echo this back-and-forth, potentially providing designers a way to articulate particular "relational shifts" when designing with/for AI (Nicenboim et al., 2024).

Ecological smallness challenges sociotechnical simplification and *invisibility*, understood here as the disappearance of technology's social and environmental traces behind APIs, dashboards or interfaces – what Latour (2005) described as the stabilization of actor-networks. Small AI replaces invisibility with *accountability*, pointing to both the (arithmetic) accounting of consequences and the resulting development of responsibility.

Parametric smallness problematizes and replaces *growth* with *sufficiency* (Santarius et al., 2023). As we note above, adding parameters to models does not guarantee better performance. In fact, smaller models are often better suited to address specific physical or cultural phenomena. Challenging excess with the logic of sufficiency opens new forms of understanding appropriateness that reject the substitution of quantity for quality and trouble the very notion that quantity can be a measure of quality.

Data smallness challenges the drive to represent, quantify and datafy the world, a drive for synoptic views that aspire to conjure a system in its *totality*. Working with smaller datasets provides specificity and an appreciation of difference. Here, 'complete' generality – 'one dataset to rule them all!' – is replaced by the prescience brought by *situatedness*, and with it, aspects of the lifeworld can be represented in ways that reflect both entanglements and specificity (see for instance Lee, Speed & Pschetz, 2024).

Distributedness counteracts the centripetal pull of large GenAI players, offering a form of resistance to the *interpellation* of individuals into the logic of corporate AI. Here *conviviality* stands as a strategic choice to reject the blandness, extractivism and absorption of the AI industry, allowing the creators and users of small LMs to enact a multitude of alternatives, each built to purpose as defined, articulated and configured by those who use them. The end goal may be seen as a form of commoning.

Lastly, *pluriversality* goes against the homogenizing imaginary of the *universal* tool, user and use. In "a world of many worlds" (Escobar, 2018) there can be no single technical apparatus that addresses everyone all the time, and insofar as technology carries forward ontological assumptions and dispositions (Heidegger, 1977), Small AI may promote a sense of unruly becoming, cultural assertion, and strivings towards *difference*.

Taken together, the relational shifts we describe (and summarize in Figure 1) provide a diagram of what it means to carry out an unmaking of AI towards smallness, and how design thinking and design acting, if undertaken from the position of smallness, may create a conceptual space for appreciating and promoting degrowth-centred design with/for AI. Each of these relational shifts can be used to frame a collection of practices that support it, to which we turn next.

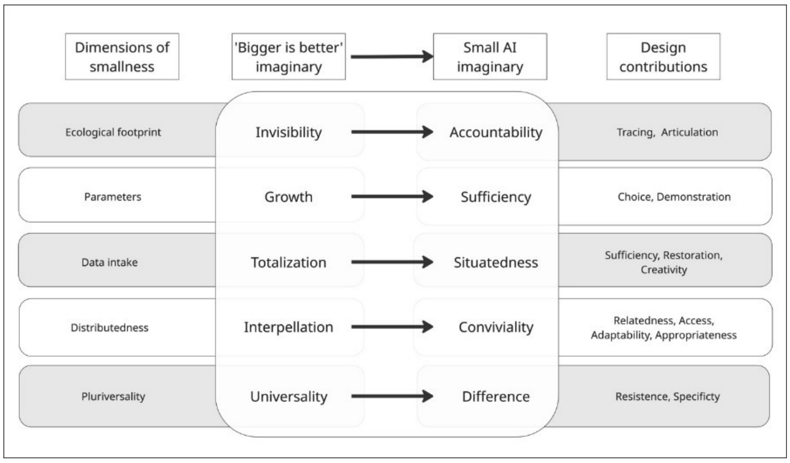

Figure 1 Overview of smallness through its relations to design practice. The Figure illustrates the relation between Small AI dimensions and the design strategies they may imply. The two are connected by what we explain here as 'unmaking' – the potential of Small AI to reconfigure practices away from the modernist, 'bigger is better' imaginary towards the humbler, Small AI imaginary.

4. An agenda for designing with/for Small AI

Because Small AI is an imaginary-in-the-making, we are reluctant to suggest rules, principles or strict guidelines for how to design with/for it. Such proposals tend to stunt the emergent potential of new practices for an illusion of order and stability. Instead, we would like to suggest a few starting points for designers to consider and act on GenAI from the position of smallness. Our suggestions are not exhaustive but are meant to gesture towards the right direction; they are not a map but a compass.

Broadly speaking, the first task of design is to amplify and sharpen the critique of mainstream, 'bigger is better' AI. Such a critique will help to counter the pull of hegemonic understandings of efficiency and resist the sense of inevitability that surrounds and promotes GenAI. A second task is to develop a better understanding of Small AI in practice, in order to create guidelines and measurements for evaluating the smallness of GenAI systems and applications. Lastly, these efforts should include communities in meaningful ways, fostering community-centred computation infrastructures and practices that respond to public needs and aspirations. In line with McQuillan (2022), we believe that the future of GenAI technologies, if it is not to promote defuturing (Fry, 2020), must incorporate significantly more community involvement and ownership.

4.1 *Ecological accountability*

While the reduction of the ecological footprint of models is largely considered the preserve of engineers and technologists, designers can make choices over the technologies they work with. They can also support the transition from a culture of invisibility to one of accountability by facilitating the consolidation of a public (DiSalvo, 2009) engaged with reducing the ecological footprint of model use. The use of *tracing*, as can be seen in projects such as Crawford and Joler's "Anatomy of an AI System" (2018), appears particularly useful, but so do other communicative and pedagogical activities aimed at *articulating* the ecological footprint of GenAI. This may include translating (Luccioni, Jernite & Strubell, 2024), labelling (Keeley & Noothout, n.d.), guiding choices (Lupetti, Cavallin & Murray-Rust, 2025), sharing best practices and suggesting metrics and standards – all necessary steps in building legibility, transparency and accountability, whose outcomes could be used to assess and adjudicate research and development projects.

4.2 Valuing parsimony

Since model parameterization is already subject to labelling – HuggingFace, for instance, lists the parameter size of each available model – the role designers can play here has more to do with promoting the value of parsimonious approaches to modelling. Designers may help undermine the inevitability floated by 'bigger is better' claims by foregrounding *choice* – intervene in community debates about optimal compute models (Hoffman et al., 2022) and highlight the fact that size does matter, that there is, in fact, choice over language models, and that useful results can be had with smaller and partly open models (such is the case with the Stable Diffusion-based app Draw Things, for instance). Designers can also help *demonstrate* the ways in which small, local models can be creatively used where the larger ones could not – whether it is because of physical embedding, responsiveness, offline use, etc.

4.3 Situating data

Questions around data sources and use are not new for those designing with/for AI (see for instance several of the chapters in Giaccardi & Bendor, 2024), however, the perspective opened by Small AI allows the following: first, in concert with engineering practices, designers can help map out answers to questions about data *sufficiency* in direct relation to generalizability, applicability and accessibility. This may require forms of intermediate level knowledge that connect the technical complexities to design criteria and desired behaviours (Höök & Löwgren, 2012). Second, designers can look at the linkages between data and the communities it represents. Along the lines of Indigenous AI practice, initiatives such as the ROAM principles – basing data usage in Rights, Openness, Accessibility and Multi-Stakeholder ways of working – illustrate how one may navigate precarious situations in *restorative* rather than disruptive ways (Llanes-Ortiz, 2023, p.47). Third, designers can look to the long history of designing by, with and for data (Speed & Oberlander, 2016) to think about ways to reduce, but also to *creatively* engage with data. In this mode, we see special value in artistic or art-inspired approaches to curating data (Arzberger, Lupetti & Giaccardi, 2024; Elwes, 2019; Hemment et al., 2024; van den Burg et al., 2023) for they suggest ways to trouble default, 'naturalized' uses of data in support of more critical, parsimonious, situated and playful tactics.

4.4 Infrastructuring conviviality

A Small AI research agenda focused on the redistribution of the power to compute may surface ways to lower the barriers for creating and using GenAI, and assessing the potentials of Small AI to promote conviviality and anchor commoning. Here we take a note from Vetter's (2018) work to imagine a distributed, anti-centripetal agenda for designing with/for Small AI – one that maintains that not all relationality is equally beneficial to communities and individuals. In Vetter's terms, *relatedness*, when applied to design with/for AI, foregrounds how models can be used to strengthen human social ties rather than human-model ties, in contrast with how GenAI vendors wield relationality – often as a form of sycophancy – as a means to retain user attention and create dependency (Kaffee, Pistilli & Jernite, 2025). *Access* points to the cost and ownership of models, but also to how they support skill building and value local knowledge. *Adaptability* speaks to the diversity, repairability and local operability of AI systems as measures of their distributive potential. And *appropriateness* points to the potential for localisation, for joyful worktime, and for the re-use of models. Together these aspects provide a possible infrastructure for convivial uses of GenAI.

4.5 Insisting on plurality

Through the perspective opened by Small AI value is regarded as a measure of *resistance* to the idea that there is a single global culture and world to model (Couldry & Mejias, 2019; Prietl, 2019), thus hinting at a potential antidote to Big Tech induced cultural hegemony and homogeneity, and inviting a move towards algorithmic pluralism (Jain et al., 2024). Designing with/for Small AI would prioritize pluriversal models, in combination with the kind of decentralized conviviality we discuss above. A pluriversal lens asks designers to interrogate the relations between plurality and commonality as they pertain to LMs, contributing to better understandings of the tradeoffs between model range and cultural *specificity*, of how to look across model diversity and value uniqueness even at the cost of generalisability, and of how to assemble collections of (small) models rather than remain tethered to monolithic services at individual and community scales. In more practical terms this means not only internalizing the fact that the world cannot be modelled as a unified entity, but also rejecting the notion that such an endeavour is valuable.

5. Conclusion

In this conceptual, agenda-setting paper we argue that the social and ecological harms of GenAI technologies necessitate both an alternative imagery and a set of concomitant practices that would resist, undermine and reconfigure the 'bigger is better' model that underlies the AI industry. We expanded the meaning of smallness from applying strictly to the ecological footprint of language models (LMs) to encompass political, epistemic and cultural dimensions. In this, we go beyond other critical approaches to AI that focus on either its ecological footprint or its social implications, but not on both, and offer a distinct designerly perspective.

We believe the notion of Small AI may help recapture the appeal of GenAI as a means for joyous exploration of possibilities. Here, a model that is smaller in aspiration might trade generalizability for precision, or reliability, or contestability. The model might be aware of the edges of its operation, especially given the tendency of language models to hallucinate or bullshit their way confidently through uncertainty (Hicks, Humphries & Slater, 2024). The text prompt may then cease to be an entry point into a large, expensive and exploitative computational apparatus, and become an invitation to an open-ended productive, reflexive and caring adventure. Designers can help in this by committing to unmaking their relations to GenAI and with that, the relations GenAI weaves together. In section 4 we made a few suggestions for what this may imply, but it is up to the design research community and, more importantly, the wider community of GenAI users to help redirect the industry's trajectory. That said, as in other cases, we should be careful not to overestimate what design can do (Lorusso, 2023) and seek a humble, yet potentially impactful position. Resisting the sense of inevitability promoted by the AI industry and contributing to an alternative imaginary for GenAI may not be sufficient but make for a good start.

6. References

Agre, P. (1997). *Computation and human experience*. Cambridge University Press.

Ahmed, N., Wahed, M., & Thompson, N. C. (2023). The growing influence of industry in AI research. *Science*, 379(6635), 884–886.

Allen, D., & Weyl, E. G. (2024). The real dangers of generative AI. *Journal of Democracy*, 35(1), 147–162.

Amoore, L., Campolo, A., Jacobsen, B., & Rella, L. (2024). A world model: On the political logics of generative AI. *Political Geography*, 113, 103134.

Arzberger, A., Lupetti, M. L., & Giaccardi, E. (2024). Reflexive Data Curation: Opportunities and Challenges for Embracing Uncertainty in Human-AI Collaboration. *ACM Transactions on Computer-Human Interaction*, *31*(6), pp.1-33.

Bai, G., Chai, Z., Ling, C., Wang, S., Lu, J., Zhang, N., Shi, T., Yu, Z., Zhu, M., Zhang, Y., Yang, C., Cheng, Y., & Zhao, L. (2024). Beyond Efficiency: A Systematic Survey of Resource-Efficient Large Language Models. *arXiv preprint*. Available: https://arxiv.org/abs/2401.00625v2

Baker, C. E. (2006). Media concentration and democracy: Why ownership matters. Cambridge University Press.

Bareis, J., & Katzenbach, C. (2022). Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values*, 47(5), 855–881.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, pp.610–623.

Bendor, R. (2025). Sustainability imaginaries by design. *Sustainability: Science, Practice and Policy*, 21(1), article 2443998.

Blythe, M., Lindley, S., & Murray-Rust, D. (2025). Artificial Intelligence and other Speculative Metaphors. *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*, pp. 347–356.

Brevini, B. (2022). *Is AI good for the planet?* Polity.

Bridle, J. (2022). Ways of being. Beyond human intelligence. Penguin.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. and Agarwal, S. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, pp.1877-1901.

Caillon, A., & Esling, P. (2021). RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. arXiv preprint. Available: https://arxiv.org/abs/2111.05011

Çelikkanat, H. (Oct.25, 2025). October 2025 Crawl Archive Now Available. *Common Crawl Blog*. Available: https://commoncrawl.org/blog/october-2025-crawl-archive-now-available

Chiang, T. (2023). ChatGPT Is a Blurry JPEG of the Web. *The New Yorker*. Available: https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web

Comber, R., & Eriksson, E. (2023). Computing as Ecocide. *LIMITS '23: Ninth Workshop on Computing within Limits*, June 14–15 2023.

Comet API (Oct.18, 2025). How Many Parameters does GPT-5 have. Available: https://www.cometapi.com/how-many-parameters-does-gpt-5-have

Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, 20(4), 336–349.

Crampton, J. (2020). Think. Resist. Act local: Is a slow AI possible? Reading Time.

Crawford, K. (2021). The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Crawford, K., & Joler, V. (2018). *Anatomy of an AI System*. Available: https://cache.fluxo.info/data/4c/ab/4cab5c66a9e44e27f58dd79918f489d7ed08da09/screenshot.pdf

De Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, 7(10), 2191–2194.

Derczynski, L. (2023). The Language Model Vocabulary Gap. *Inter human agreement*. Available: https://interhumanagreement.substack.com/p/demons

DiSalvo, C. (2009). Design and the Construction of Publics. *Design Issues*, 25(1), 48–63.

Dignum, V. (2019). Responsible artificial intelligence: How to develop and use AI in a responsible way. Springer.

Dove, G., & Fayard, A. L. (2020). Monsters, metaphors, and machine learning. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20),* pp. 1-17.

Engelbutzeder, P., Borning, A., Ballmaier, A., & Wulf, V. (2025). Small is Still Beautiful. Towards a Reorientation of HCI and Computing. *International Reports on Socio-Informatics (IRSI),* 22(4), 3-45.

Elwes, J. (2019). Zizi. Available: https://www.jakeelwes.com/project-zizi-2019.html

Escobar, A. (2018). Designs for the pluriverse: Radical interdependence, autonomy, and the making of worlds. Duke University Press.

Feola, G. (2019). Degrowth and the Unmaking of Capitalism: Beyond "Decolonization of the Imaginary"? *ACME: An International Journal for Critical Geographies*, 18(4), 977–997.

Frauenberger, C. (2019). Entanglement HCI the next wave? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 27(1), 1–27.

Fry, T. (2020). Defuturing: A new design philosophy. Bloomsbury.

Fuchs, C. (2015). Social media and the public sphere. In *Culture and economy in the age of social media* (pp. 315–372). Routledge.

Fouquet, R. & Hippe, R. (2022). Twin Transitions of Decarbonisation and Digitalisation: A Historical Perspective on Energy and Information in European Economies. *Energy Research & Social Science* 91: 102736.

Giaccardi, E. & Bendor, R. (eds., 2024). *Rethink Design: A Vocabulary for Designing with AI*. TU Delft Open.

Girdhar, N., Raj, A., Sharma, D., Singh, V., Doucet, A. & Renz, M. (2025). A Comprehensive Review of Frugal Artificial Intelligence: Challenges, Applications, and the Road to Sustainable AI. *Soft Computing* 29(13–14): 4823–56.

Heidegger, M. (1977). *The question concerning technology, and other essays* (W. Lovitt, Trans.). Harper & Row.

Hickel, J. (2023). On Technology and Degrowth. *Monthly Review*, 75(3). Available: https://monthlyreview.org/2023/07/01/on-technology-and-degrowth

Hicks, M. T., Humphries, J. & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26(2), 1–10.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., de Las Casas, D., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., van den Driessche, G., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., … Sifre, L. (2022). An empirical analysis of compute-optimal large language model training. *Advances in Neural Information Processing Systems*, 35, 30016–30030.

Höök, K., & Löwgren, J. (2012). Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Transactions on Computer-Human Interaction*, 19(3), 23:1-23:18.

Illich, I. (1973). *Tools for Conviviality*. Calder and Boyars.

Istrate, R., Tulus, V., Grass, R. N., Vanbever, L., Stark, W. J., & Guillén-Gosálbez, G. (2024). The environmental sustainability of digital content consumption. *Nature Communications*, 15(1), 3724.

Jain, S., Suriyakumar, V., Creel, K., & Wilson, A. (2024). Algorithmic pluralism: A structural approach to equal opportunity. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FACT)*, 197–206.

Jones, J. C. (2021). *Designing Designing*. Bloomsbury.

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 12(6), 518–527.

Kaffee, L.-A., Pistilli, G., & Jernite, Y. (2025). INTIMA: A Benchmark for Human-AI Companionship Behavior. arXiv preprint. Available: https://doi.org/10.48550/arXiv.2508.09998

Keeley, I. & Noothout, L. (n.d.). AI scoring beyond technical performance. *Clever Franke*. Available: https://www.cleverfranke.com/research/ai-scoring-beyond-performance

Khanal, S., Zhang, H., & Taeihagh, A. (2025). Why and how is the power of big tech increasing in the policy process? The case of generative AI. *Policy and Society*, 44(1), 52–69.

Kwet, M. (2024). *Digital Degrowth: Technology in the Age of Survival*. Pluto Press.

Latour, B. (2005). Reassembling the social: an introduction to actor-network-theory. Oxford University Press.

Lee, Y., Speed, C. & Pschetz, L. (2024). Pheno-data: Using tomatoes to rethink data and data practice for ecological worlds. *Human–Computer Interaction*, 40(1-4), pp.249-271.

Lifset, R., Porter, A., Newman, N. & Lee, T. (2025). Artificial Intelligence’s Environmental Impacts: Exploring Research on “Red AI”. Available at SSRN: https://ssrn.com/abstract=5311903

Lin, J., Zhu, L., Chen, W.-M., Wang, W.-C., & Han, S. (2023). Tiny machine learning: Progress and futures. *IEEE Circuits and Systems Magazine*, 23(3), 8–34.

Lindley, J., Akmal, H. A., Pilling, F., & Coulton, P. (2020). Researching AI Legibility through Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.

Lindström, K., & Ståhl, Å. (2020). Un/Making in the Aftermath of Design. *Proceedings of the 16th Participatory Design Conference,* Volume 1, pp.12–21.

Lindström, K., & Ståhl, Å. (2023). Un/Making the Plastic Straw: Designerly Inquiries into Disposability. *Design and Culture*, 15(3), 393–415.

Llanes-Ortiz, G. (2023). *Digital initiatives for indigenous languages*. UNESCO. Available: https://unesdoc.unesco.org/ark:/48223/pf0000387186

Lorusso, S. (2023). *What Design Can't Do*. Set Margins.

Luccioni, S., Jernite, Y., & Strubell, E. (2024). Power hungry processing: Watts driving the cost of AI deployment? *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FACT)*, pp. 85–99.

Lunden, I. (Feb.11, 2025). AI investments surged 62% to $110B in 2024 while startup funding overall declined 12%. *TechCrunch*. Available: https://techcrunch.com/2025/02/11/ai-investments-surged-62-to-110-billion-in-2024-while-startup-funding-overall-declined-12-says-dealroom

Lupetti, M. L., & Murray-Rust, D. (2024). (Un)making AI Magic: A Design Taxonomy. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, pp. 1–21.

Lupetti, M. L., Cavallin, E., & Murray-Rust, D. (2025). The Unbearable Lightness of Prompting. Critical Reflection on Sustainability and genAI use in Design Education. arXiv preprint. Available: https://arxiv.org/abs/2501.16061

Machado de Oliveira, V. & Cinnamon Tea, A. (2025). *Standing in the Fire: A Speculative Inquiry into Meta-Relationality and Generative AI.* Available: https://decolonialfuturesnet.wordpress.com/wp-content/uploads/2025/08/standing-in-the-fire-interim-report-2025.pdf

McQuillan, D. (2022). Resisting AI. An anti-fascist approach to artificial intelligence. Bristol University Press.

Mistral AI (July 22, 2025). Our contribution to a global environmental standard for AI. Available: https://mistral.ai/news/our-contribution-to-a-global-environmental-standard-for-ai

Murray-Rust, D., Nicenboim, I., & Lockton, D. (2022). Metaphors for designers working with AI. *Proceedings of the Design Research Society's Biennial Conference (DRS '22)*, pp. 1-19.

Nenno, S. (2024). Potentials and limitations of active learning: For the reduction of energy consumption during model training. *Weizenbaum Journal of the Digital Society*, 4(1).

Nicenboim, I., Søndergaard, M. L. J., Lindley, J., Reddy, A., Strengers, Y., Redström, J., & Giaccardi, E. (2024). Unmaking-with AI: Tactics for Decentering through Design. *ACM Trans. Comput.-Hum. Interact.*, 31(6), 82:1–82:20.

Nolan, E. (2021). Metaphoring back in the climate crisis: Notes toward encouraging student engineer agency through metaphoring. *Working Papers in Applied Linguistics and Linguistics at York*, 1.

O'Brien, I. (Sept.15, 2024). Data center emissions probably 662% higher than big tech claims. Can it keep up the ruse? *The Guardian*. Available: https://www.theguardian.com/technology/2024/sep/15/data-center-gas-emissions-tech

Pasquinelli, M. (2023). The eye of the master: A social history of artificial intelligence. Verso Books.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). Carbon Emissions and Large Neural Network Training. arXiv preprint. Available: https://www.kathimerini.gr/wp-content/uploads/2024/07/2104-10350_1.pdf

Pinhanez, C., Cavalin, P., Storto, L., Finbow, T., Cobbinah, A., Nogima, J., Vasconcelos, M., Domingues, P., Mizukami, P. de S., Grell, N., Gongora, M., & Gonçalves, I. (2024). Harnessing the Power of Artificial Intelligence to Vitalize Endangered Indigenous Languages: Technologies and Experiences. arXiv preprint. Available: https://doi.org/10.48550/arXiv.2407.12620

Prietl, B. (2019). Big data: Inequality by design? *Weizenbaum Conference*, 10.

Rakova, B., & Dobbe, R. (2023). Algorithms as social-ecological-technological systems: An environmental justice lens on algorithmic audits. *arXiv* preprint. Available: https://arxiv.org/abs/2305.05733

Rehak, R. (2025). AI Narrative Breakdown. A Critical Assessment of Power and Promise. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 1250–1260.

Rehak, R. (2026) Catastrophic Computation. On the Impossibility of Sustainable Artificial Intelligence. In: Hagedorn, L., Schmid, U., Winter, S., Woltran, S. (eds.) *Digital Humanism*. Springer, Cham, pp. 110-118.

Rehak R., Ullrich, A., & Vladova, G. (2025). Contesting Openness in AI. Reflections on the transformative value and participatory potential of open source AI. *Proceedings of the 2025 ACM International Conference on Information Technology for Social Good (GoodIT '25)*, pp. 217-225.

Rohde, F., Wagner, J., Meyer, A., Reinhard, P., Voss, M., Petschow, U., & Mollen, A. (2024). Broadening the perspective for sustainable artificial intelligence: Sustainability criteria and indicators for artificial intelligence systems. *Current Opinion in Environmental Sustainability*, 66, 101411.

Sabie, S., Jackson, S. J., Ju, W., & Parikh, T. (2022). Unmaking as Agonism: Using Participatory Design with Youth to Surface Difference in an Intergenerational Urban Context. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1-16.

Sabie, S., Soden, R., Jackson, S., & Parikh, T. (2023). Unmaking as Emancipation: Lessons and Reflections from Luddism. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1-15.

Santarius, T., Bieser, J. C., Frick, V., Höjer, M., Gossen, M., Hilty, L. M., ... & Lange, S. (2023). Digital sufficiency: conceptual considerations for ICTs on a finite planet. *Annals of Telecommunications*, 78(5), 277-295.

Schumacher, E. F. (1973). Small is beautiful; economics as if people mattered. Harper & Row.

Schütze, P. (2024). The problem of sustainable AI: A critical assessment of an emerging phenomenon. *Weizenbaum Journal of the Digital Society*, 4(1).

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63.

Smith, M. (Oct.1, 2025). The Hidden Behemoth Behind Every AI Answer. *IEEE Spectrum*. Available: https://spectrum.ieee.org/ai-energy-use

Speed, C., & Oberlander, J. (2016). Designing from, with and by Data: Introducing the ablative framework. *Proceedings of the Design Research Society Biennial Conference (DRS '16)*. pp. 2991-3004.

Steig, F., Koenig, P. D., Marquardt, J., Oels, A., Radtke, J., Rehak, R., & Weiland, S. (2025) Sustainability powered by digitalization? (Re-)politicizing the debate. *Sustainability: Science, Practice and Policy*, 21(1), article 2521181.

Sourati, Z., Ziabari, A.S., & Dehghani, M. (2026). The homogenizing effect of large language models on human expression and thought. *Trends in Cognitive Science*. Article 2834.

van den Burg, V., Boer, G. de, Akdag Salah, A. A., Chandrasegaran, S., & Lloyd, P. (2023). Objective Portrait: A practice-based inquiry to explore AI as a reflective design partner. *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, 387–400.

Varoquaux, G., Luccioni, A. S., & Whittaker, M. (2025). Hype, sustainability, and the price of the bigger-is-better paradigm in AI. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FACT)*. pp.61-75.

Vetter, A. (2018). The Matrix of Convivial Technology – Assessing technologies for degrowth. *Journal of Cleaner Production*, 197, 1778–1786.

Virz, M. & Huang, V.G. (Feb.2, 2025). AI Is Wall Street's Modern Gold Rush. Here's How Investors Are Placing Their Bets. *Wall Street Journal*. Available: https://www.wsj.com/tech/ai/ai-wall-street-investments-companies-0baba8d9

Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J., & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7–30.

Widder, D. G., Whittaker, M., & West, S. M. (2024). Why "open" AI systems are actually closed, and why this matters. *Nature*, 635(8040), 827–833.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the frontier of power. Profile Affairs.

About the Authors:

Roy Bendor is Associate Professor of Critical Design and Director of the Design for Interaction Master's programme at Delft University of Technology. His research explores the capacity of design to disclose and promote alternative social, political and environmental futures.

Dave Murray-Rust is an Associate Professor at TU Delft and co-director of the AI Futures Lab. He blends artistic and creative practice with empirical and theoretical research, exploring what happens when humans and technology become entangled.

Rainer Rehak is a researcher at the Weizenbaum Institute and the Berlin Social Science Center. His work combines computer and social science, ethics and epistemics, digitalization and sustainability, and convivial digital de-colonialism. He co-initiated the "Bits & Bäume" conferences for digitalization and sustainability.