2. Integrating feminist theories into design: the case of participatory Artificial Intelligence

Stefana Broadbent

In an effort to address the climate crisis that is transforming our world, social science disciplines are engaged in a systematic scrutiny of the economic and social models that have so rapidly led to the current environmental degradation. Among the analytic frameworks being deployed, feminist theories are providing a longstanding tradition of critical analysis of the dynamics of power, exploitation and dualistic thinking. These concepts are particularly relevant when wanting to understand the common roots underlying the exploitation of natural resources and social and gender inequality. The urgency of adopting a renewed perspective on society calls for both theoretical and practical approaches, and the solutions which take into account the gender dimension are increasingly seen as a fruitful avenue to address both the expressions of structural inequality and of natural resource depletion. Starting from this objective, recent eco-feminist theories propose an alternative framing of resource management and conservation as an integrative ecology of just economies within living worlds. Eco-feminist proposals for climate justice and environmental preservation are based on local, decentralized, pluralistic economic and social models that are inclusive, just and potentially regenerative. Feminist systemic models of environmental and social response to the crisis offer the field of design a set of ideas that can well fit into the traditions of participation and social innovation. Feminist design is emerging as an alternative voice that can bring together the different streams of design for social justice, environmentalism, policy, and postcolonialism. In 2024, for instance, two books have been published exploring the potential intersection between design and feminism, *Feminist Designer: on the Personal and the Political in Design*, an edited volume by Alison Place (2023), which collects the writings of more than fourty designers to examine how to innovate the design process; and *Designing Gender: a feminist toolkit* by Elsie Baker (2023), which explores design projects which challenge gender inequality. The focus of both books is not on the role of women in design, but how feminist theories can inform design processes and projects.

The integration of machine learning and artificial intelligence in most domains where design operates, make them a central locus of reflection for designers and may offer the opportunity to use the eco-feminist lens to expand the boundaries of intervention. Feminist perspectives on AI are raising the issues of material and labour exploitation in the *production* of AI systems, and discrimination and victimisation in the consequences of their *application* (Eubanks, 2018). While AI is being hailed as a potential avenue for addressing the climate crisis, feminist scholars are joining their voices to those of other critics who are raising concerns about the devastating material costs in terms of energy and water consumption involved in running these systems (Crawford, 2021); on the indiscriminate appropriation of data produced by humans for private profit (Couldry and Mejas, 2019) and the risks of profound injustice in the application of algorithmic models in decision-making processes that concern people's lives (Hildebrandt, 2021).

Viewed from any perspective, AI technologies in their current mode of development seem to be predicated and entrenched in a logic of exploitation, and as eco-feminist scholars argue, are the product of a worldview in which extraction of value can come at the expense of certain natural categories. However, alternative approaches are emerging which centre on ideas of participatory AI, democratic AI and distributed AI. These are all attempts to bring AI into spaces of public governance and democratic values, and to anchor the impact of potential benefits emerging from these technologies to the construction of common goods. They also are directions that fully resonate with social innovation models while offering frameworks that would allow practices to progress towards more systemic transformations.

2.1 Feminist theories and eco-feminisms

Eco-feminism lies at the intersection between two primary political movements of the late 20th and early 21st centuries: feminism and environmentalism. Eco-feminist theories strive to expose the interdependence of social inequalities and environmental consequences, and conversely the effect of environmental degradation on the increase of marginalisation (Warren, 1997). The eco-feminist critique (Gaard, 2010) bases itself on viewing global capitalism as a patriarchal structure based on the exploitation of, not only women, but of the colonised, the poor and the non-human environment (fauna, flora, and ecosystems in general). The growth in global capitalism, beginning with the Industrial Revolution and massively expanded after the Second World War, has brought about huge technological, economic, and scientific advancements, but is inextricably entrenched in the unprecedented abuse of nature and peoples. The feminist perspective points out the western dualistic models of rationality that distinguish men from nature (Plumwood, 1993) and attributes a hierarchy of domination and subordination to each of them, as a root cause. Within this worldview, nature is considered irrational, unpredictable, potentially hostile (all categories attributed equally to women, who are seen as being closer to nature) and therefore requiring to be dominated and controlled through rationality which is defined as a superior category of thought. In this context, to be classified as *nature* means to be defined as a passive resource, a background with limited agency, available to be used and moulded, and *naturally* supposed to be dominated.

In the feminist critique, the step from considering everything in nature inferior and needing to be controlled, to exploitation and extractivism, has led to economic models that are extremely destructive for the environment and unjust for women and minorities. According to Jessica Weir (2009), thinking hyper-separation «places humans in a relation of mastery with respect to earth others and limits their capacity to respond to ecological devastation. Humankind loses the ability to empathise and see the non-human sphere in ethical terms». The realm of nature in other words, becomes an unlimited resource and a receptacle for waste. Thus, in this analysis, the current capitalist economic system is based on the extraction of resources, be they in the natural or the social world, with little or no reciprocity. The exploitation includes not only the existing natural resources such as minerals, oil, water, land, but also the control of future resources through the managed reproduction of fauna and flora to serve economic purposes, and more recently, human experiences through the collection of data.

2.2 Critical theories of data and AI

From the vantage point of feminism, but also of science and technology studies, systematic analyses on data and artificial intelligence have emerged that raise serious concerns as to the implications of the current frameworks within which large AI systems are being developed, namely: extreme centralisation, private ownership, deregulation, and appropriation of resources be they natural or social.
These are the same concerns, incidentally, that underlie the regulatory efforts in the EU and US (Halim and Gasser, 2024).
To describe such economic models, Couldry and Meijas (2019) have coined the term *colonialist machine learning*. Kate Crawford (2021) has pointed out the materiality of the extractivist model underlying AI development; Timnit Gebru (2021) has criticised the significant biases in data models; and Eubanks (2018) provides some damning examples of the injustice arising from the application of its models to vulnerable populations.

The two books by Couldry and Mejias (2019 and 2024) make the strong claim that the extraction of data from peoples' online activity by companies working in the digital industries, is a direct continuation of colonialism and the exploitation of natural resources by colonising economies in previous centuries. Their argument is not metaphorical, they insist that data exploitation, as appropriation of human life through data, is a new form of exploitation. They show that digital platforms in all domains from work to leisure health and education, capture and translate our lives into data, and then extract information that is fed into enterprises which then sell it back to us. This dispossession of human experience happens because these human features are just there, free to be taken, devalued, exactly like indigenous land was *just there* to be claimed. Critique of data grabbing has been voiced by many other commentators of digital economy (Zuboff, 2018; O'Neil, 2017; Acemoglu, 2021; Mazzucato *et al*., 2022; Broadbent *et al*., 2024) and is the current object of lawsuits and regulatory efforts such as the EU AI Act of 2024.

Kate Crawford (2021) on the other hand, starts by reminding us of the materiality of digital economies. Our digital services are made of cables, devices, server farms, all of which are made of minerals that must be mined, plastics that are difficult to recycle, and all of which require vast amounts of energy and water to run (Cara, 2022). Current numbers show that there are 4 billion devices in the world, 1.4 million km of underwater cables, and 8000 data centres. The materiality of the digital ecosystem raises the well-known issues of resource depletion, pollution, and carbon production. The CO_2 impact of digital services is currently estimated at 2.1% to 3.9% of overall emissions (Freitag *et al*., 2021), which is higher than the airline industry. AI is expected in its current form to increase these numbers even more and have a devastating impact on climate change by contributing to the transformation of the soil through mining, the pollution of fresh water, and the burning of fossil fuel for energy production.

Moving to issues of labour, digital services and AI rely on hundreds of thousands of low-wage workers, usually in emerging economies, to moderate content (Casilli, 2019), train machine-learning algorithms, correct and improve outputs from systems and maintain the infrastructure. The labour laws that regulate these jobs are under considerable scrutiny as the level of precariousness is exceedingly high.

Finally, in the list of issues that digital services and AI, in particular, are raising, is the question of bias and injustice. There is a vast amount of literature on the biases that are inbred in the databases and data sources used by large machine learning systems (Gebru, 2021; O'Neil, 2017; Acemoglu, 2021), which has shown the consequences of skewed sampling and underrepresentation of some populations. Eubanks, 2018 has given a chilling account of the biases inbuilt into the algorithms and the real-life consequences that predictive systems can bring about with weighing variables or performing stepwise model selection on datasets. The Feminist Generative AI Lab, started by the Convergence program AI, Data & Digitalisation and led by Sara Colombo at TU Delft, is also challenging standard practices in data science, which can perpetuate and reinforce existing biases and power imbalances.

The growing field of feminist critique of AI (Wajman, 2021; Browne *et al.*, 2023; Noble, 2018; Nissenbaum, 2021), highlights many of the same issues that have been raised in analysing the social and ecological consequences of traditional industries. The centralsation of production, distribution and governance which characterises the energy industries, for instance, defies principles of social justice by reducing participation in the decision-making processes, the distribution of benefits and costs and representation of the people and entities concerned. In the field of machine learning, Browne (2023) exposes the structural injustice of predictive systems that are left in the hands of private ownership, and the limitations of a traditional regulatory approach which focusses on liability limitation. Browne pushes for a new form of public body with citizen representation capable of bringing to the table the contextual and underlying dynamics of structural injustice.

> **On substantive questions such as how personal data ought to be collected and how its use be governed, or how much analysis should be done on the biased outcomes of algorithms before their assessments and predictions become the bases of policy, or how ought the Government to plan to counter the socio-economic effects of automation of certain labour market tasks, it is highly likely that a group of citizens would draw substantially different conclusions to those of industry experts or politicians. I argue that this is the key to creating a very different sort of public-body**

approach to AI-generated structural injustice than the models we currently have in play (Browne, 2023, p. 365).

The damning analysis of the ecological impact of AI technology exposed by eco-feminist and climate activists (Crawford, 2021; Cara, 2022; Monserrate, 2022) pushes the question of civic control and participation in the development of these technologies also in the realm of governing the infrastructures that enable them. One such example is the requirements of the huge server farms where computations are executed. Their consumption of energy and fresh water are so impactful on the localities where they are implanted, that local governance is paramount to avoid an unfair distribution of resources between citizens and digital companies. In communities where 30% of the energy and water risks being directed to the data centres, the decision can *only* be collective and democratic.

Considering all the complex issues mentioned above, there is an emerging consensus that participatory AI is the only way to avoid the perpetuation of the structural problems of the economic models of the last 50 years, potentially at a far greater scale given the expected impact of machine learning technology on society and the environment.

2.3 Designing participatory AI

In recent years Nesta's Centre for Collective Intelligence has started analysing and funding projects that attempt to integrate collective intelligence and machines. Crowdsourcing information and knowledge, as has been done by the most successful examples of collectively created knowledge commons such as Wikipedia, OpenStreetMap and Linux, is a complex endeavour that requires content and governance models to be managed. On the other hand, citizen platforms such as Decidim (deployed in Barcelona, Reykjavik, Helsinki etc.) to collect opinions and suggestions; the participatory budgeting systems tested in many EU cities, or the wide consultations such as the EU Conference for the Future of Europe, have produced vast quantities of citizen-generated content that have been costly to analyse and synthesise. The scientific projects such as Zooniverse, which have been a reference in the field of citizen science, have also hit some hurdles in maintaining the commitment and participation of their volunteers, as have the patient groups that harness knowledge from its members (Broadbent, 2014; Nesta, 2015).

Nesta has therefore started looking at ways to integrate AI systems in collective intelligence projects and has contributed to the funding of some initiatives that attempt to see collectives designing machine learning systems to improve their work (Nesta, 2021).

One of these projects is Sepsis Watch from the Duke Institute for Health Innovation, a sepsis detection and management platform that uses deep learning to predict the likelihood of a patient developing sepsis. The Sepsis Watch model was trained to identify cases based on dozens of variables. Its training data consisted of 50,000 patient records with more than 32 million data points. It was successfully integrated into hospital operations, with data flowing from electronic patient records and alerts being incorporated into physicians' workflows. The original proposal to develop an AI-based solution was driven by a team of frontline doctors. The team included implementation experts, machine learning experts, and clinical experts. Participatory design was used to improve the accuracy and appropriateness of the technology solution and importantly, to retain agency and control of decision-making for clinical staff.

The first twelve months of the project were used to establish the team, characterise the problem, and start designing the data pipeline and work- flow for the model. First of all, clinical experts curated the local datasets and selected the parameters that the model was trained on. After this, the teams dedicated one year to developing the AI system, and integrating it into a user-facing platform which became Sepsis Watch. After a model was created, clinicians evaluated the performance of the model based on known cases of sepsis, which led to further fine-tuning. Together with nurses, the clinical experts also reviewed multiple versions of the user interface for the tool (Nesta, 2021).

Another participatory model of AI development that Nesta describes (2021) is a project in the Mazvihwa Communal Area, Zimbabwe, where land management problems were arising from woodland grazing areas being transformed in cropland. The Muande Trust, a local community research organisation, helped develop the Zimbabwe Agro-Pastoral Management Model to explore potential systemic behaviours under a variety of rainfall variation scenarios and combinations of management interventions. Using participatory modelling, local stakeholders helped define the parameters and data to be used by the model and examined the impact of different types of interventions through simulations. The model visualised different actions and impacts which led to question some land management practices and led to policy changes that allowed the reuse of fallow fields for farming (Eitzel *et al*., 2020).

As in the case of Sepsis Watch, the crucial element of participation was in the definition of the model itself and in particular in the balance given to persistence over time rather than average annual harvest.

> **We defined persistence as a set minimum amount of cows, woodland, and harvest at the end of every model year; we calculated average annualized harvest by dividing total accumulated harvest by the number of years before the modelled system dropped below any of the persistence thresholds (if it did so). Average annualized harvest was therefore a shorter-term measure of sustainability: a particular run could maximize harvest at the expense of livestock numbers or woodland biomass and only last a few years but with potentially excellent harvest, resulting in a value of "not persistent" and a high annual harvest for that run.**
> **In contrast, persistence was a longer-term measure of sustainability: a model run might last all 60 calendar years with cows, crops, and woodland above the persistence thresholds, while the average harvest over that time might be correspondingly lower (representing a classic resilience trade-off). From a climate adaptation sovereignty perspective, the people of Mazvihwa should define their own persistence thresholds: what constitutes "enough" harvest, cows, or woodland for a village the size of Mudhomori (approximately 100 households in 2013). (Eitzel *et al*., 2020, p. 7).**

These brief examples show that it is possible within complex AI systems to envisage design processes where stakeholders govern the definition of goals, of data and algorithms (by defining and weighing the significant variables), and the testing and evaluation of the outcomes. In both cases local distributed participatory models drove the technology to produce benefits for the community concerned.

While participatory models of technology development are certainly not new, the complex and opaque nature of machine learning technology raises new challenges in terms of enabling greater control by multiple stakeholders. The Collective Intelligence Project, an organisation which aims to create better and more collectively-intelligent models of governing transformative technologies such as AI, has proposed a framework for Democratic AI. At the core of the approach is the idea that it is possible to develop more processes for public input into AI systems and manage the collective governance of training data to improve the data supply chain, including opt-out and transparency processes. A governance model that is squarely in the hands of stakeholders and the public is presented as a way of ensuring that infrastructures, design and implementation, impacts and oversight are in the public domain. This approach not only shifts the control of the technology but also promotes it for projects of public interest. The issues related to the environmental impact of such technologies are therefore subsumed within a strategy of public good. It is possible to imagine, as proposed by legislators and technologists, that choices of deployment would be made, also taking into consideration their environmental effects.

2.4 Conclusions

The longstanding tradition of human-centred and participatory design has been a first step towards integrating the worldviews of people involved in transformations of their physical or social environments. The feminist perspective adds a layer that is often missing in human-centred approaches – the issue of power and environmental degradation. While this question is set in the broad context of structural injustice, it does lead to pushing the boundaries of participation beyond individual agency and the ability to carry out goals and intentions. Formulating an innovative approach to AI development that transfers the control of data, development implementation and infrastructure in the public realm, is a way of extending participatory approaches to encompass a more significant control over resources. The model of decentralised, localised, just and pluralistic forms of management proposed by feminist theories can constitute a roadmap for expanding the scope of participatory design. Embedding design in the work of defining ownership, governance, monitoring and legislation as well as the structures of interaction between the different actors, means embracing systemic transformations.

References

Acemoglu D. (2021), "Harms of AI", *NBER Working Papers 29247*, National Bureau of Economic Research, Cambridge.

Baker S. E. (2023), *Designing Gender: A Feminist Toolkit,* Bloomsbury Publishing, London.

Bell S. E., Daggett C., and Labusky C. (2020), "Toward feminist energy systems: Why adding women and solar panels is not enough", *Energy research & social science*, 68.

Browne J., Cave S., Drage E. and McInerney K., eds. (2023), *Feminist AI: critical perspectives on algorithms, data, and intelligent machines,* Oxford University Press, Oxford.

Broadbent S. (2015), *Collective Intelligence: questioning the individual approach to skill development,* in Dolphin T. ed., *Technology, globalisation and the future of work,* IPPR, London.

Broadbent S., Khamass M., Forestier F. and Zolinsky C. (2024), *Pour une nouvelle culture de l'attention,* Odile Jacob, Paris.

Buckley C. (2020), "Made in patriarchy II: Researching (or re-searching) women and design", *Design Issues*, 36, 1: 19-29.

Cara F. (2022), *Il digitale sostenibile. Gli obiettivi ambientali e sociali,* in Acanfora M. edited by, *Ecologia digitale. Per una tecnologia al servizio di persone, società e ambiente,* Altra Economia, Milano.

Casilli A. (2019), *En attendant les robots. Enquête sur le travail du clic,* Le Seuil, Paris.

Cooke B. and Kothari U. (2001), *Participation: the new tyranny?*, Zed Books, London.

Couldry N. and Mejias U. A. (2019), *The Costs of connection. How data are colonizing human life and appropriating it for capitalism*, Stanford University Press, Redwood.

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

Dourish P., Lawrence C., Leong T. W. and Wadley G. (2020), "On being iterated: The affective demands of design participation", *Proceedings of the 2020*

CHI Conference on Human Factors in Computing Systems April 25-30, 2020, Honolulu, USA.

Eitzel M. V., J. Solera K. B., Wilson K., Neves A. C., Fisher A., Veski O. E., Omoju A., Mawere N. and Mhike Hove E. (2020), "Indigenous climate adaptation sovereignty in a Zimbabwean agro-pastoral system: exploring definitions of sustainability success using a participatory agentbased model", *Ecology and Society,* 25, 4: 13-23.

Eubanks V. (2018), *Automating inequality: How high-tech tools profile, police, and punish the poor,* Picador St. Martin's Press, New York.

Feenstra M. and Özerol G. (2021), "Energy justice as a search light for gender energy nexus: Towards a conceptual framework", *Renewable Sustainable Energy Reviews*, 138.

Freitag C., Berners-Lee M., Widdicks K., Knowles B., Blair G. S. and Friday A. (2021), "The real climate and transformative impact of ICT: a critique of estimates, trends, and regulations", *Patterns*, 2,9.

Gaard G. (2010), *Ecofeminism*, Temple University Press, Philadelphia.

Gaard G. (2015), "Ecofeminism and climate change", *Women's Studies International Forum,* 49: 20-33.

Gebru T., Morgenstern J., Vecchione B., Vaughan J. W., Wallach H., Iii H. D. and Crawfod K. (2021), "Datasheets for datasets", *Communications of the ACM*, 64, 12: 86-92.

Halim N.L. and Gasser U. (2023), *Vectors of AI Governance - Juxtaposing the U.S. Algorithmic Accountability Act of 2022 with The EU Artificial Intelligence Act.* Available at SSRN: https://ssrn.com/ abstract=4476167

Hildebrandt M. (2021), "Discrimination, data-driven ai systems and practical reason", *European Data Protection Law Review*, 7, 3: 358-366.

Introna L. and Nissenbaum H. (2000), "Shaping the web: Why the politics of search engines matters", *The Information Society*, 16, 3: 169-185.

Kern L. (2021), *Feminist city: Claiming space in a man-made world*, Verso Books, New York.

Mazzucato M., Schaake M., Krier S. and Entsminger J. (2022), "Governing artificial intelligence in the public interest", *UCL Institute for Innovation and Public Purpose, Working Paper Series* 2022, 12.

Mejias U. A. and Couldry N. (2024), *Data grab: The new colonialism of big tech and how to fight back*, University of Chicago Press, Chicago.

Monserrate S. G. (2022), "The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage", *MIT Case Studies in Social and Ethical Responsibilities of Computing,* Winter 2022.

Nesta (2015), *Collective Intelligence in patient organisations,* Nicholas L. and Broadbent S. eds. Available at https://www.nesta.org.uk/report/col-lective-intelligence-in-patient-organisations/

Nesta (2021), *Participatory AI for humanitarian innovation: a briefing paper.* Berditchevskaia A., Malliaraki E., Peach K., eds. Available at https://www.nesta.org.uk/report/participatory-ai-humanitarian-innovation- briefing-paper/

Noble S. U. (2018), *Algorithms of oppression: How search engines reinforce racism*, New York University Press, New York.

O'Neil C. (2016), *Weapons of math destruction: How big data increases inequality and threatens democracy*, Crown Publishing Group, New York.

Place A. ed. (2023), *Feminist Designer: On the Personal and the Political in Design,* The MIT Press, Cambridge.

Plumwood V. (1993), *Feminism and the Mastery of Nature*, Routledge, New York.

Srnicek N. (2017), *Platform capitalism*, Polity Press, Cambridge.

Toupin S. (2024), “Shaping feminist artificial intelligence”. *New Media & Society*, 26, 1: 580-595.

Wajcman J. (2020), *Pressed for time: The acceleration of life in digital capitalism*, University of Chicago Press, Chicago.

Warren K. J. ed. (1997), *Ecofeminism: Women, culture, nature,* Indiana University Press, Bloomington.

Warren K. J. (2000), *Ecofeminist Philosophy: A Western Perspective on What It Is and Why It Matters,* Rowman & Littlefield, Lanham.

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