



# Shaping feminist artificial intelligence

new media & society  
2024, Vol. 26(1) 580–595  
© The Author(s) 2023



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/14614448221150776  
journals.sagepub.com/home/nms



**Sophie Toupin** 

Concordia University, Canada

## Abstract

This article examines the historical and contemporary shaping of feminist artificial intelligence (FAI). It begins by looking at the microhistory of FAI through the writings of Alison Adam and her graduate students to enrich the plural histories of AI and to write back feminist history into AI. Then, to explore contemporary examples of how FAI is being shaped today and how it deploys a multiplicity of meanings, I provide the following typology: FAI (1) as model, (2) as design, (3) as policy, (4) as culture, (5) as discourse, and (6) as science. This typology sheds light on the following questions: What does the term FAI mean? How has FAI been shaped over time?

## Keywords

Algorithms, Alison Adam, civil society, critical AI, datasets, feminist AI, feminist history, gender, race

## Introduction

The first time I came across the term *feminist artificial intelligence* (FAI) I was doing research on rule-based feminist chatbots. I was looking for examples of chatbots that were using AI to “automate feminism” as I was about to present some preliminary thoughts on feminist chatbots at the 2020 Algorithms for Her conference in London, UK. At the time, I did not find any feminist chatbots powered by a feminist data set and trained by feminists to do FAI. Instead, what I found were online presentations and a handful of projects that used and described the term in varying ways. Josie Young’s (2019) “Why We Need to Design Feminist AI” Ted Talk on YouTube argued for the necessity of FAI to counter the sexist and harmful effects of AI-powered voice assistants like Siri and Alexa. Charlotte Webb’s (2019) “What Is a Feminist AI: Possible Feminisms,

---

## Corresponding author:

Sophie Toupin, Concordia University, 1515 Saint-Catherine Street W, Montreal, QC H3G 2W1, Canada.  
Email: [sophie.toupin@mail.mcgill.ca](mailto:sophie.toupin@mail.mcgill.ca)

Possible Internets” at the Disruption Network Lab in Berlin suggested ten guiding principles building on the work of feminist Human Computer Interaction (HCI) scholar Shaowen Bardzell (2010). The projects included the Los Angeles community-based research and design group called *feminist.ai* which used the software Poieto to involve unheard voices in AI design and Caroline Sinderson’s (2020) artistic project to create a feminist data set through a series of public workshops with the goal of powering a FAI system. This is the broad context that prompted me to ask: What does the term FAI mean? How has FAI been shaped over time?

The category of *feminist artificial intelligence* reflects the ways in which “feminism” and “AI” have come to take on a multiplicity of meanings. While I use the term FAI in the singular to reflect how the term is used today, I recognize that this analytical category is plural due to the contemporary manifestations of feminisms including their intersectional (Combahee river collective, 1979; Crenshaw, 1991), Black feminist (hooks, 2015), decolonial (Lugones, 2010), and liberal inflections. Similarly, the term AI is used in the singular and may take different meanings depending on who is using the term. It may refer to the usage in the 1955 paper *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (McCarthy et al., 1955), where AI was defined as an attempt “to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves” (p. 1). It may be used to refer to the general idea that AI is an ideology that serves the doctrine of efficiency (Katz, 2020). Or it may be understood as a generic all-encompassing signifier that describes the growing computer network and infrastructure that rely on big data and algorithms for predictions. Finally, it is important to note that most scholars agree that General AI does not yet exist meaning anything to do with sentient robots. What is currently being referred to as AI is narrow AI meaning statistics on steroid (Broussard, 2018). It works by “analyzing an existing dataset, identifying patterns and probabilities in that dataset, and codifying these patterns and probabilities into a computational construct called a model” (p. 32). In this article, I do not take a fixed definition of either AI or feminism and rather use them in the ways in which the scholars and practitioners being discussed have mobilized these terms.

Methodologically, this article is not based on a systematic and exhaustive review of all material published on the subject. Rather it is a critical survey of some writings and practices on the subject that shed light on the different understandings of the broad category of FAI. It is a purposive sample that I have found in academic databases (Google Scholar and WorldCat), in search engines (Google Chrome and Mozilla Firefox), and on Twitter, where a lot of these scholars and activists are active. I primarily took into consideration the scholarly work and practices of those who use the broad category *feminist artificial intelligence* (or FAI), which includes among other designations intersectional AI (Ciston, 2019), posthuman FAI (Meinders, 2017), transfeminist AI (Peña and Varon, 2020), FAI from below (West, 2020), and intersectional decolonial AI (Costanza-Schock, 2018). While I have consulted articles and websites in English, French, Portuguese, and Spanish, the bulk of the articles, books, and practices referenced in this article and which currently exist are in English. Finally, I have made a considerable effort to look for practices and writings from the global South to broaden our imaginaries of who is involved

in crafting this notion, and to show how the term has been generative in a variety of contexts and for different purposes. This is why I have gone beyond North American and European research and practices on FAI and spotlighted individuals, collectives, projects, and research located in the majority world. Shedding light on undervalued knowledge production within AI is a feminist move that allows scholars, readers and practitioners to have access to alternative narratives and to enrich the history and contemporary practices of critical AI research and practice.

This article is an exploration of past and current efforts to think through the notion of FAI and how this practice challenges traditional and/or hegemonic forms of AI. First, I give a micro-historical overview of the context in which FAI has emerged and how it came to be. This historical section enriches an understanding of AI, showing that feminist scholars have long been thinking about, articulating, and prototyping FAI. Further, it makes visible the contradictions and questions that the authors have revealed and struggled with when prototyping their small-scale projects. Then, to explore contemporary examples of how FAI is being shaped today, I provide the following typology: FAI (1) as model, (2) as design, (3) as policy, (4) as culture, (5) as discourse, and (6) as science. Finally, I conclude by making the case that while feminist techies, activists, and artists are critically engaged in understanding and doing AI differently, they have yet to engage beyond the technoliberal sphere of inclusion.

## The historical shaping of FAI

Feminist critiques of AI are rooted in the turn toward constructivist perspectives within the philosophy and sociology of science and within feminist science and technology studies (FSTS). In the 1980s, this body of scholarship started arguing that Western conversations on science were incomplete and skewed because women's knowledge was excluded (Jansen, 1988; Turkle, 1984). Later, this field would open to other excluded forms of knowledge including from the majority world, indigenous and non-humans (Harding, 2008; Liboiron, 2021; TallBear, 2013). FSTS focused on breaking or reconciling the boundaries between nature and animal, humans and machines, and mind and body in addition to questioning objectivity in science and critiquing the long-standing EuroAmerican hegemonic masculinity within technoculture (Forsythe, 1993; Wajcman, 1991). The posture of situated knowledge (Haraway, 1988) was a breakthrough in FSTS as it foregrounded the idea that knowledge production is situated in social relations, that all knowledge derives from partial perspectives, and that perspectives taken from subjugated positions provide the most "objective" accounts of the social worlds from which they emerge. Influenced by these conversations, British computer scientist and historian of science Alison Adam started criticizing AI from a feminist perspective showing it had deeply conservative foundations (Suchman, 1987, 2008). Her aim was to bring together gender, race, and class specifically and power more generally into the critique of AI by asking: How is AI used and for which purposes? How does AI represent knowledge? And what knowledge is used in AI systems? (Adam, 1995). These feminist questions constituted her starting point to then formulate the notion of FAI, not as an "alternative successor to AI . . . [but rather] how AI can be informed by feminist theory and be used for feminist projects" (Adam, 1998: 157).

This section briefly tells the microhistory of FAI through Adam's work and relationships. It makes the case that her scholarship deserves attention because many of her questions are still relevant today and most contemporary practices cite her as an inspiration. In the context in which Adam was thinking about FAI there was a profound disenchantment with AI research. Despite considerable public investment in the 1970s and 1980s, AI had not yet been able to deliver on its promises (Jansen, 2002; Mitchell, 2019). Projects she stressed "do not attract research funding and must often be tackled within the confines of final year undergraduate and masters (MSc) level dissertations" (Adam, 1998: 158). There was a clear frustration from Adam that few funding bodies and universities were interested in this FAI practice and hence it did not flourish outside of her lab.

The oldest reference I found to the term "feminist artificial intelligence" appears in the conclusion of Adam's (1995) article *Artificial Intelligence and Women's Knowledge*. In this article, Adam asserted that the way forward for a different AI informed by feminism is twofold. First, it is possible to use traditional AI for feminist purposes. Second, feminism needs to become involved in AI. It was, however, in her book *Artificial Knowing* that Adam (1998) furthered her thoughts about FAI. But before explaining what she meant by the term, she raised three conceptual concerns, all drawing from the work of critical feminist scholars that came before her (Jansen, 1988; Suchman, 1987; Turkle, 1984). First, she believed that AI (meaning the object and the people studying and developing it) had controversial epistemological and philosophical understandings of what constitutes intelligence. Intrinsic to the question of intelligence is the problematic model of rationality of AI projects. In the context of AI research and practice, the term "rationality" is used to suggest that the human mind has precedence over the body and fails to account for the culturally and socially situated nature of intelligence; in short, its thinking and usage privileges a disembodied form of AI. With the emergence of robotics, however, Adam acknowledged that the physical embodiment of AI had figured prominently, but the cultural embodiment of AI had not. Her third conceptual concern was AI and male domination. She specifically proposed to undo the rational "male as norm" idea in AI, as the dominant paradigm of AI at the time relied on knowledge that pretended to be unsituated, or what Haraway (1988) called the "god trick" of seeing everything from nowhere.

To explain what she meant by FAI, Adam heavily relied on the work of two of her graduate students who prototyped under her guidance feminist knowledge-based systems. Knowledge-based systems, better known in symbolic AI parlance as expert systems, were a prevalent AI model at the time. The first FAI she described was prototyped in 1993 by her Brazilian student Chloe Furnival, who is currently a faculty member in computer science at the Federal University of São Carlos in Brazil. The FAI project aimed to compile feminist jurisprudence on issues of gender discrimination in the United Kingdom. This prototype was programmed in PROLOG and was premised on the fact of the burden of proof in gender discrimination cases resting on women to show their rights had been violated. The system relied on a feminist jurisprudence database to support low-income women pursuing gender discrimination cases who did not have access to a lawyer to make a compelling case in front of a judge. Reflecting on the challenges with the development of this prototype, Adam and Furnival (Adam, 1998; Adam and Furnival, 1995) asked the following: How do we design an AI-based system that would not be

threatening for the low-income female user who has never used a computer before? How do we take into consideration gender, race, and class in the design and the outcome? And how do we not give too much hope in relation to the given advice, nor push a decision onto the user? This feminist legal knowledge-based system was never implemented in real life because of financial constraints.

The second FAI project undertaken by Adam via her British graduate student Maureen Scott was a feminist computational language knowledge-based system. The goal of the system was to repair in a meaningful way misunderstandings in communication between men and women. The system started from the premise that women are usually the agent of repair in gendered conversations. Wanting to value women's repair model of communication, the feminist knowledge-based system made them the experts in this field. The knowledge necessary for making this prototype was acquired through conversations with feminists. Again, this system never went beyond a student thesis project.

Looking back at these two experiments, Adam stressed that these systems were full of unresolved contradictions. They were in fact replicating some of her own critique of AI, especially when it came to the disembodied aspect of knowledge inputs in the system and the decision that came out as a result. Her reflexive process led her to ask what she had identified as the dominant and weak engineering technology question: "I had to start somewhere" (Adam, 1998: 166). Paraphrasing Audre Lorde's statement that you can't use the master's tools to demolish the master's house, Adam (1998) believed that feminist AI could be used tactically to "make a few scratches in the masonry" (p. 70). In fact, she stated, "there are as yet no other tools and we cannot know until we try" (Adam, 1998: 164), bringing to the fore the need to move away from only criticizing AI, and instead move toward doing FAI (Adam, 1998; Adam and Furnival, 1995). This question of doing and making has been a recent theme in FSTS: "[m]oving beyond critique to practice is the next critical move" (Bauchspies and Puig de la Bellacasa, 2009: 341).

Recognizing some of the dangers associated with doing FAI, Adam told the story of being asked by a man in a gender and technology workshop how a fighter jet plane designed by a feminist would look different. She replied that "feminists do not design fighter planes" (Adam, 1998: 157). This statement alludes to the political commitment associated with technology, implying that certain forms of FAI should not be developed. A second aspect associated with this story was her uncertainty as to how different the two feminist knowledge-based systems were from traditional expert systems. Her worry was that both the feminist legal system and the linguistic experiment ended up replicating the social order and power structures (Adam, 1998) that exist in society. In other words, this means that they were part of what Atanasoski and Vora (2019) call the technoliberal approach to solving social, cultural, political, and economic problems. Notwithstanding, when answering the question of whether feminists should design AI applications, she replied that despite the existing contradictions we can't lose sight of the political project of FAI.

## Current trends shaping FAI

Why is FAI re-emerging? First, FAI is resurfacing within an "AI spring" meaning an increased availability of funding for research and projects. The massive funding being

allocated to deep learning followed Geoffrey Hinton's team win at the 2012 ImageNet competition that showed supervised machine learning was extremely effective at "predictive pattern recognition when trained using significant computational power and massive amounts of labeled data" (Whittaker, 2021: 51). Second, critical technology scholars have shown that the increasing use of predictive analytics is tightly embedded within racial capitalism, and it has deleterious impact on equity deserving groups (Benjamin, 2019b; Crawford, 2021). Black feminist scholars for instance have revealed how Black girls and women have been negatively impacted by algorithms and other systems (Benjamin, 2019a; Buolamwini and Gebru, 2018; Noble, 2018). This scholarly work has been essential to set the stage for the development of a critical turn in AI research and prototyping, which includes among other elements decolonial/decolonizing AI (Adams, 2021; Mohamed et al., 2020), Indigenous AI (Lewis et al, 2020), and FAI (West, 2020).

In the following section, I make visible the people, processes, relationships, and technologies that are necessary for FAI to exist. I show that just like Adam and her students' prototypes, FAI projects remain small scale, incomplete, and full of contradictions. To better comprehend what contemporary FAI means, I propose the following typology: FAI as model, design, policy, discourse, culture, and science. These categories are at times porous and capacious, but all in all they help us expand our understanding of what FAI is and does.

## **FAI as model**

The first typology I explore is FAI as model. "Model" here refers to what Broussard (2018) defines as "a kind of black box that we can feed data into and get an answer out of. We can take the model and run new data through it to get a numerical answer that predicts something" (p.32). A model depends on a data set to be improved and fine-tuned. The discussion regarding the historical shaping of AI, above, focused on two small-scale knowledge-based prototypes that demonstrate one example of FAI as model. Below, I focus on one small-scale contemporary FAI project that ultimately aims to use a machine-learning model as a counterpoint to large-scale corporate-led opaque and extractive deep learning systems.

The machine-learning-design researcher and artist Caroline Sindors (2020) exemplified well what contemporary FAI as model means. Her Feminist Data Set project launched in 2017 examined every step of the AI process in the hope of developing a small-scale community-led data set. The goal of this data set is ultimately to feed an FAI-powered chatbot with whom a user can converse on feminisms. The project entails breaking down the main tasks required for an AI-powered chatbot to function: data collection, data labeling, data training, selecting the algorithm model to use, as well as designing how the model is placed in a chatbot and what the chatbot looks like (Sindors, 2020). Through a series of in-person workshops and forums held in public spaces such as art galleries and bookstores in London (UK), Sindors invited feminists to collaborate on the process of creating a data set composed of pieces of work that spoke to the multiplicity of feminisms. The open process she followed was inspired by Wikipedia's peer production approach, meaning that the data was created by a community that cared enough for it to exist. This made visible the invisible work that has been historically

performed by women and often today by people from the majority world (Posada, 2022; Roberts, 2021). It was also an educational project for the community to understand why the data in data sets matter in AI systems. As Sinders remarked, the process of data collection needed to be slow to address gaps and correct bias. After every workshop, Sinders carefully looked at the data collected and asked the following: Are there too many works by cisgender women or white women in the data set that will skew the conversation with the chatbot? If the answer was yes, she specifically called for pieces of work from non-cis women, women of color, and trans creators to enrich her data set (Sinders, 2020). Taking time to collect the data with a feminist community that responded to her call was a key part of creating the data set on which her chatbot would rely to converse on feminisms.

The second step Sinders took to develop her chatbot pertained to the data labeling process. She decided to prototype the technically responsible knowledge (TRK) tool to make visible the tedious and underpaid labor of labeling the data. This tool included a wage calculator designed to provide a necessary living wage for feminist data laborers and labelers, showing that this type of work is important and needs to be valued in FAI development. With the design of this tool, she offered a feminist alternative to the exploitative Amazon Mechanical Turk platform. Sinders' FAI-powered chatbot is still in the making, showing how slow and time-intensive these projects are. Notwithstanding, the feminist social and political commentary that is coming out of this project offers an alternative path for doing AI. In fact, Sinders (2020) has framed her ongoing project as an act of protest, because she has been facilitating the invisible, slow, and nitty-gritty work of creating a community-led data set, a practice that goes against the efficiency, productivist, black box and unbridled pace of large-scale corporate and government machine-learning projects. The Feminist Data Set project is a window into understanding the slow, small-scale, incomplete artistic and educational nature of FAI as model and the need for feminist data sets.

## **FAI as design**

The most popular way of understanding FAI is without a doubt through the lens of design. This reflects an orientation of AI as engineering technology, meaning that it is about design and the building of computer systems. Already in the 1990s Adam had argued that the professionalization of AI designers and the need for prestige had given legitimacy to understanding AI as an engineering technology over other meanings. This allowed "AI researchers to concentrate on the design process and the building of artifacts and to leave [aside] more difficult philosophical questions [such as] the nature of intelligence" (Adam, 1998: 14). (Trans) feminist scholars as well have been increasingly interested in doing and designing feminist science and technology including feminist HCI (Bardzell, 2010), design justice (Costanza-Shock, 2018) and data feminism (D'Ignazio and Klein, 2020) among others. Below, I use one concrete example to illustrate what FAI as design may mean by focusing on cultural aspects, and the deleterious impact of corporate feminism.

The art and design research group called *feminist.ai* is based in Los Angeles. Started in 2016, this group was created as part of Christine Meinders' graduate work where she

organized community workshops to provide participants with access to AI design tools (Meinders, 2017; Meinders et al., 2020). Meinders' approach to FAI stems from Adam's recognition that the cultural embodiment in AI design is lacking. To facilitate the integration of culture as a key design consideration in engineering AI, Meinders developed a software called Poieto. "Culture" in her design process refers as much to the culture of the makers of AI as it does to the culture of the users and to the greater societal context in which AI tools are developed. The culture of the makers means that she attends to increasing the participation of women, queer, trans, and non-binary designers from all backgrounds, origins, and geographies. Behind this articulation is the assumption that the gender, race, and other identities of the designers will have an impact on the development of AI. The culture of the users means that participatory design with women, queer, trans, and Black, Indigenous and People of Color (BIPOC) individuals is part and parcel of Poieto. At the heart of this design feature is the aspiration to lower the barriers to entry for doing AI for the community to be engaged in this process, despite their low to no coding experience. The promise of including community in the design of AI is believed to diminish the adverse impact of these systems on the community. The culture of the societal context in which AI tools are developed means that certain groups are negatively and disproportionately impacted by AI systems that are developed by corporations with huge capital.

The unequal and contradictory tech Californian environment in which the founder and members of *feminist.ai* find themselves led Meinders to trademark the expression "Feminist AI™." Registering this expression as part of the United States Patent and Trademark Office gives Meinders the right to use it in the United States and internationally for renewable periods of 10 years. Using this method of intellectual property common among corporations, Feminist AI™ becomes a brand to be protected rather than an expression to be used as part of the digital commons. *Feminist.ai* as a small-scale project sits uneasily between a corporate type of feminism and a more community-oriented type of organization.

## FAI as policy

The category of FAI as policy comes to life in the context of the push for developing countries to adopt and catch up with the AI race, and for women not to be left out from the assumed economic benefits. FAI becomes an international development policy issue in the settler colonial context of Canada. In 2015, when liberal Prime Minister Justin Trudeau came into power, he implemented a Feminist International Assistance Policy which promotes gender equality, inclusion, and the empowerment of women and girls including in science, technology, engineering, and mathematics (STEM). This policy is strengthened by Canada's growing interest in responsible AI, and a renewed commitment to the United Nations Sustainable Development Goals (SDGs). In 2018, the International Development Research Center (IDRC), a Canadian federal Crown corporation that funds research and innovation as part of Canada's foreign affairs and development efforts, published a white paper on AI and development in the global South. This spearheaded a series of call for proposals to fund artificial intelligence for development (AI4D). As a result, the Feminist AI Research Network ( $f < a + i > r$ ) was born.

Initiated by the Swiss NGO Women at the Table and their global South university partners from Costa Rica and Thailand, the  $f < a + i > r$  received funding from IDRC. The 3-year (2021–2024) project valued at \$2 million CAD (IDRC, 2021) aims to incubate FAI. Called Feminist AI: From Paper to Prototype to Pilot this project represents the first initiative of its kind to substantially fund FAI. The project presents itself as intending to correct the current corporate logic in which AI systems are developed in the North and then exported to the South alongside the export of systemic gender, racial, and intersectional bias. With this pool of money,  $f < a + i > r$  funds digital rights NGOs which center women and LGBTQI+ people from the global South as key actors and developers of FAI. Their wager is that global South NGOs will help debias data and improve AI systems.

As of December 2022,  $f < a + i > r$  has selected a few papers to go to a prototyping phase. The first prototype being built is from the Argentinian Foundation Via Libre. This organization wants to prototype a tool that aims to reduce discriminatory behaviors within automatic processing of language through word embeddings for text. A good example of discriminatory word embedding is with terms like “nurse,” “care,” or “sewing” usually being associated with women, whereas terms like “salary,” “sports,” and “leader” are associated with men (Alemany et al, 2022). The second prototype is built by the NGO DataGenero, an observatory of data with a gender perspective also based in Argentina. The project consists of an open-source AI system that will collaborate with criminal court officials in Argentina and Mexico to bring more transparency to the judicial treatment of gender-based violence (GBV) cases against women and LGBTQI+ people. The goal of the system is “to increase the general public’s trust in justice” and for governments to “widely adopt evidence-based public policies that prevent GBV and eradicate it before these cases may become feminicides” (Feldfeber et. al, 2022: 20). The third prototype is by the Chilean NGO Derechos Digitales. Their project aims to develop a practical guide for the collaborative design of FAI systems. Derechos Digitales stresses that the guide will be based on local needs and concerns and be developed by those who will be impacted by these systems (Guerra, 2022).

FAI as policy is a story about settler colonial Canada that wants to become a champion of global “responsible AI” by supporting a feminist liberal perspective on science and technology. This is expressed through the realization of SDG 5 (Gender Equality) in increasing women in the field of STEM, and SDG 16 (Peace, Justice and Strong Institutions) to promote AI and good practices by states and other institutions in the global South. What is however silenced from FAI as policy in a time of environmental duress is the fact that mining companies headquartered in settler colonial Canada are involved in extraction of minerals on indigenous land abroad. It will be necessary to follow closely how exactly these projects will unfold and who they will in the end benefit to be able to grasp what FAI as policy has to offer. The question I am left with is the following: How has FAI come to be seen as a responsible or appropriate technology for the development of gender equality and as a response to GBV? What role is the availability of funding AI projects playing in shifting the work of feminist techies and digital rights organizations toward FAI?

## FAI as discourse

Until now, I have focused on categories of FAI that are more practice-based than discursive. In this section, I examine FAI as discourse as a legitimate category to understand the types of imaginaries associated with this notion.

FAI is often used as a placeholder or signifier to describe critical work that addresses the relationship between gender and technology, and at times, at the intersection of race and class. The first example I use is the women's and gender studies department at the University of Florida, which uses the rubric "Feminist AI and Big Data" to refer to academic books such as Alison Adam's *Artificial Knowing*, Safiya Umoja Noble's (2018) *Algorithms of Oppression*, and Ruha Benjamin's (2019a) *Race after Technology*. This use of FAI as discourse signals to the reader that feminist, queer, and critical race theories are being used to examine and critique AI systems.

The second example I use is the Radical AI Podcast and its episode on FAI 101. In this episode, the hosts of the podcast interviewed two researchers on the meaning of FAI. Considering FAI as a feminist technology that is doing good in the world, Dr Kerry Mackereth gave the example of the multimedia AI-powered chatbot called Mumkin (Doyle-Burke and Smith, 2021). This Indian feminist app was designed to have difficult conversations to address the issue of female khatna (also known as female genital cutting) within the Bohra community in India. The chatbot is intended to help users shed inhibitions and gain confidence to have real-life conversations on gender, culture, and society. Acknowledging the difficulty in defining FAI, Dr Mackereth preferred to explain what it is not: "FAI will never be engaged with predictive policing" (Doyle-Burke and Smith, 2021).

FAI has a discursive power that is effectively used by practitioners and advocates. Renata Avila (2021), who is involved with  $f < a + i > r$ , has suggested the need to reframe AI along feminist social justice principles to unleash new AI imaginaries. She wrote that FAI

reframe[s] the discourse around AI, and brings it closer to the hands and minds of women and girls [. . .] Feminists will not just translate today into the design, they we'll translate utopia and code gender and race equality into the design. (Avila, 2021)

Similarly, Neema Iyer (n.d.), the director of the Kampala-based NGO Policy, has spoken of FAI as discourse in an African context to refer to how feminist movements in Africa can critically engage, discuss, and envision a feminist future under different AI and data governance models. In these two examples, the discourse around FAI creates a window for women and girls to get interested and active in AI. This is likely one of the reasons why the discourse around FAI remains so positive among practitioners and scholars alike—hopeful and at times even utopian. In fact, just like AI, FAI makes grand claims about an anticipated, but yet to be realized, emancipatory potential. The question I am left with is: to which extent the current discourse around FAI renders AI palatable to feminists who might otherwise have refused it as a response to social, political, environmental and economic injustices?

Conceptually, Sarah Myers West (2020) offers us a glimpse into how to think differently about FAI as discourse. She has suggested the need for shifting the discourse around AI from fairness, accountability, and transparency to talking about remediation when imagining FAI from below (West, 2020). While West's move is discursive, it is also material, since she believes that discourse goes hand in hand with social transformation. West has criticized the liberal framework of making algorithms unbiased and inclusive, and has argued that remediation offers a powerful alternative. Remediation is understood as the work of making things right, of stopping and reversing the damage done by harmful AI systems. Fairness, she believes, distracts from the more difficult work of grappling with the systemic and structural dimensions of discrimination (West, 2020). This shift toward algorithmic remediation, she has argued, might allow for spaces of possibility and in turn give way to collective power, solidarity, and refusal (West, 2020).

## FAI as culture

Representing science and technology as culture is not new. Feminists have long framed technology as culture. In her book *Feminism Confronts Technology*, Judy Wajcman (1991) has addressed the relationship between technology and culture by articulating the mutual shaping of technology and gender relations. She wrote that “women’s exclusion from, and rejection of, technology is made more explicable by an analysis of technology as culture that expresses and consolidates relations among men. Technologies bear the imprint of the people and social context in which they developed” (Wajcman, 1991: 22). Wajcman (1991) argued that technology was recognized as masculine culture as a result of the historical and cultural construction of gender. It was an ideological position, the ideology of masculinity which created a bond between men and machines. In her discussion about technology and masculinity, she wrote that “[t]reating technology as a culture has enabled us to see the way in which technology is expressive of masculinity and how, in turn, men characteristically view themselves in relation to machines” (Wajcman, 1991: 149). She has also insisted that there are a lot of cultural factors that might lower the interest of girls and women in technology such as its associated symbols, images, and representations.

Starting from the principle that gender and race bias exist in some AI systems, Galit Wellner and Tiran Rothman (2020) took a cultural norm perspective to FAI. They classified gender biases into three categories of norms—social, physical, and cognitive—to understand what they are, where they come from, and ultimately how to correct them (Wellner and Rothman, 2020). They understood social norms as the association of a given technology with a gender (such as a dishwasher designed with women in mind), physical norms as taking one gendered body as universal (such as facial recognition systems that disproportionately recognize white men over black women), and cognitive norms as the perceived mental skills and abilities that have an impact on your life (such as cognitive racial bias, in which a postal code determines one’s approval for a loan). Changing these cultural norms, the authors believed, would lead to FAI. In practice, FAI as culture comes to life with the aforementioned software Poeto. This cultural AI tool is based on the feminist.ai research group’s 15 philosophical

orientations, one of which specifically state that culture, material and purpose are just as important as the data and model.

## **FAI as science**

The last category under study is FAI as science which focuses on research into “intelligence” from feminist perspectives. Over the years, an exploration of the meanings of intelligence within AI has emanated from disciplines such as philosophy and psychology. The only reference found that treats FAI as science in the sense that the term “intelligence” is questioned comes from the book *Surrogate Humanity* by Neda Atanasoski and Kalindi Vora (2019). Their last chapter, “On Technoliberal Desire, Or Why There Is No Such Thing as a Feminist AI,” makes their argument clear. Like Adam, they argued that the history of AI has been rationalist and masculinist from its inception, but instead of concluding that FAI is still worthy, they concluded that it comes from a technoliberal desire to be inclusive and needs a complete overhaul. They defined technoliberalism as “the ideology that technology advances human freedom and postracial futurity by asserting a postlabor world in which racial difference, along with all human social difference, is transcended” (Atanasoski and Vora, 2019: 28). To make their case, they analyzed an AI called Luna through the lens of feminist intelligence. Luna is a robot hologram created in 2018 by the humanitarian not-for-profit organization Robots Without Borders. The holographic robot is portrayed as a woman with blue eyes and dark long hair. When the robot speaks, she spells out content drawn from Wikipedia among other sources. Luna was designed to increase the quality of education and provide medical advice in humanitarian settings. Atanasoski and Vora criticized the Luna AI project for being part of a technoliberal imaginary that plays into, first, the AI marketing hype where AI is believed to be a technological solution to socio-economic problems and, second, the inclusive representation card reaffirming gendered and racial logics without the hard work needed to challenge historical and contemporary forms of intersectional oppression. This AI project is in fact quite different from the small-scale FAI projects under study since it does not see itself as being part of the FAI movement. Nonetheless, Atanasoski and Vora (2019) philosophical question that challenge the existence of a feminist intelligence is conceptually rich. They wrote,

Desire for the expansiveness of the category of intelligence, rather than a desire to disrupt this category and others that constitute the liberal subject, will not redress the surrogate effect of artificial intelligence. If we define feminism as a decolonizing project, instead of a liberal inclusive one, such that feminism politically seeks to disrupt the categories of use, property, and self-possession rather than redress through inclusion, then perhaps we might provocatively say that there need not be such a thing as a feminist intelligence (2019: 196).

Atanasoski and Vora’s provocation questions the promise of a better future that comes with the notion of FAI. AI, be it feminist or not, follows the digitization of societies. FAI tend to blind us to the dependencies we, as societies, have adopted for more

efficiency, accuracy, and convenience. FAI as it stands is less about human freedom, a postracial and postlabour future; rather, it remains as part of the extractivist capitalist system based on racial inequality, dispossession, exploitation, and the destruction of the earth. What might be a feminist decolonial (artificial) intelligence then if it is not an inclusive liberal project? Might part of the starting point lie in the recognition that rather than trying to imprison “feminist intelligence” in algorithms, data, and AI, our sophistication lies with the feminist decolonial know-how of resisting and reimagining the world?

## Conclusion

This article has explored past and current efforts to think through the notion of FAI and how these practices attempt to challenge to varying degrees traditional and hegemonic forms of AI. Looking back at the history of FAI through the writings of Adam and her students allows us to enrich the plural histories of AI and is a way to write feminist history back into AI. This microhistory has an important lesson for the present since it has not shied away from acknowledging the contradictions that exist in doing FAI.

Currently, the contemporary efforts to develop FAI have been to include diverse actors in the making of AI as well as involving communities into small-scale, slow data and AI projects. What we have come to understand is that FAI projects today are being shaped by feminist academics, students, NGOs, and artists in the global South and North. The fact that such actors are getting involved in rethinking AI brings a different flavor from the conventional history of AI. The rise of digital rights NGOs and the professionalization of their feminist activist employees has led to the involvement of new actors who study, engage with, and prototype FAI.

The promise associated with FAI is that a fairer, slower, consensual, collaborative AI is possible. The belief goes that if FAI was the default model, the types of technologies we build, and experience would be very different than the ones we have now. Outside of such utopian imaginary, FAI represents a tactical intervention and can be considered as a form of resistance to large-scale hegemonic and discriminatory AI.

FAI is however limited in what it can accomplish, as it has yet to engage beyond the technoliberal sphere of inclusion. FAI does little to challenge the power and inequalities of the capitalist political-economic system we live in that destroys the environment. FAI is not a transformative form of politics that will help change the world under extreme environmental and injustice duress because it stays within the realm of the technical fix. Technologies including data and AI can support and amplify feminist activism, but we should not forget that what we need is to further build movements in order to struggle for our collective liberation.

## Acknowledgements

I would like to thank all of those who contributed to my thinking on feminist AI including Beth Coleman, Alex Ketchum, Yousif Hassan, Fenwick McKelvey and Alessandra Renzi. I would also like to thank the peer reviewers for their comments.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Sophie Toupin  <https://orcid.org/0000-0001-7639-8180>

## References

- Adam A (1995) Artificial intelligence and women's knowledge. *Women's Studies International Forum* 18(4): 407–415.
- Adam A (1998) *Artificial Knowing: Gender and the Thinking Machine*. London and New York: Taylor & Francis Group. Available at: <http://ebookcentral.proquest.com/lib/mcgill/detail.action?docID=165123>
- Adam A and Furnival C (1995) Designing intersections—designing subjectivity: feminist theory and praxis in a sex discrimination legislation system. *Information & Communications Technology Law* 4(2): 161–173.
- Adams R (2021) Can Artificial Intelligence be decolonized? *Interdisciplinary Science Reviews* 46(1–2): 176–197.
- Alemay LL, Benotti L, González L, et al (2022) A tool to overcome technical barriers for bias assessment in human language technologies. Available at: <https://drive.google.com/file/d/1Ppdf6K8p4rYv5U-MLUql4o7fwTpoKheW/view>
- Atanasoski N and Vora K (2019) *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Durham, NC: Duke University Press.
- Avila R (2021) Deploying Feminist AI. *Alliance +*, 13 September. Available at: <https://feministai.pubpub.org/pub/deploying-feminist-ai/release/1>
- Bardzell S (2010) Feminist HCI: Taking stock and outlining an agenda for design. In: *CHI '10: Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1301–1310. Available at: <https://doi.org/10.1145/1753326.1753521>
- Bauchspies WK and Puig de la Bellacasa M (2009) Feminist science and technology studies: a patchwork of moving subjectivities. An interview with Geoffrey Bowker, Sandra Harding, Anne Marie Mol, Susan Leigh Star and Banu Subramaniam. *Subjectivity* 28: 334–344.
- Benjamin R (2019a) *Race after Technology: Abolitionist Tools for the New Jim Code*. London: Polity Press.
- Benjamin R (ed.) (2019b) *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life*. Durham, NC: Duke University Press.
- Broussard M (2018) *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, MA: MIT Press.
- Buolamwini J and Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, PMLR 81: 77–91.
- Ciston S (2019) Intersectional AI is essential: polyvocal, multimodal, experimental methods to save artificial intelligence. *Journal of Science and Technology of the Arts* 11(2): 3–8.
- Combahee river collective (1979) A black feminist statement. *Off Our Backs* 9(6): 6–8.
- Costanza-Chock S (2018) *Design Justice, A.I., and Escape from the Matrix of Domination*. Available at: <https://doi.org/10.21428/96c8d426>
- Crawford K (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

- Crenshaw K (1991) Mapping the margins : intersectionality, identity politics, and violence against women of color. *Stanford Law Review* 43(6): 1241–1299.
- D'Ignazio C and Klein LF (2020) *Data Feminism*. Cambridge, MA: MIT Press.
- Doyle-Burke D and Smith J (2021) Feminist AI. *The Radical AI Podcast*, 2 June. Available at: <https://radicalai.podbean.com/e/feminist-ai-101-with-eleanor-drage-and-kerry-machereth/>
- Feldfeber I, Quiroga YB, Guevara C, et al (2022) Feminisms in Artificial Intelligence: Automation tools towards a feminist judiciary reform in Argentina and Mexico. Available at: <https://drive.google.com/file/d/1VKeeSdJz6-8DoEbetQ8f1wTH1ORqvIXV/view>
- Forsythe DE (1993) Engineering knowledge: the construction of knowledge in artificial intelligence. *Social Studies of Science* 23(3): 445–477.
- Guerra J (2022) Towards a feminist framework for AI development: From principles to practice. Available at: [https://www.derechosdigitales.org/wp-content/uploads/Fair\\_Doc\\_Eng.pdf](https://www.derechosdigitales.org/wp-content/uploads/Fair_Doc_Eng.pdf)
- hooks b (2015) *Feminism Is for Everybody : Passionate Politics* (2nd ed., Vol. 11 online resource). London: Routledge.
- Haraway D (1988) Situated knowledges : the science question in feminism and the privilege of partial perspective. *Feminist Studies* 14(3): 575–599.
- Harding SG (2008) *Sciences from below: Feminisms, Postcolonialities, and Modernities*. Durham, NC: Duke University Press.
- IDRC (2018) Artificial intelligence and human development: Towards a research agenda. Available at: <https://www.idrc.ca/en/stories/artificial-intelligence-and-human-development> (accessed 20 April 2022).
- IDRC (2021) Advancing research on feminist artificial intelligence to advance gender equality and inclusion. *IDRC*. Available at: <https://www.idrc.ca/en/project/advancing-research-feminist-artificial-intelligence-advance-gender-equality-and-inclusion>
- Iyer N (n.d.) What does the future of Feminist AI look like? *Goethe Institute*. Available at: <https://www.goethe.de/prj/k40/en/eth/fem.html>
- Jansen SC (1988) The ghost in the machine: Artificial Intelligence and gendered thought patterns. *Resources for Feminist Research/ Documentation sur la Recherche Feministe* 17(4): 4–7.
- Jansen SC (2002) *Critical Communication Theory: Power, Media, Gender and Technology*. Lanham, MD: Rowman & Littlefield.
- Katz Y (2020) *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York: Columbia University Press.
- Lewis JE, Abdilla A, Arista N, et al. (2020) Indigenous protocol and Artificial Intelligence position paper. Available at: <https://spectrum.library.concordia.ca/id/eprint/986506/>
- Liboiron M (2021) *Pollution Is Colonialism*. Durham, NC: Duke University Press.
- Lugones M (2010) Toward a decolonial feminism. *Hypatia* 25(4): 742–759.
- McCarthy J, Minsky ML, Rochester N, et al. (1955) A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine* 27(4): 12.
- Meinders C (2017) From Feminist AI projects to artificial knowing. *Media Design Practices Lab*, 30 November. Available at: <https://mediadesignpractices.net/lab/2017/11/collaborative-ai-research/>
- Meinders C, Thompson J, Ciston S, et al. (2020) Week 3 : Feminist AI (Main Thread). *Critical Code Studies*. Available at: <http://wg20.criticalcodestudies.com/index.php?p=/discussion/87/week-3-feminist-ai-main-thread>
- Mitchell M (2019) *Artificial Intelligence : A Guide for Thinking Humans*. 1st ed. New York: Farrar, Straus and Giroux.
- Mohamed S, Png MT and Isaac W (2020) Decolonial AI: decolonial theory as sociotechnical foresight in Artificial Intelligence. *Philosophy & Technology* 33(4): 659–684.

- Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
- Peña P and Varon J (2020) Decolonising AI : A transfeminist approach to data and social justice. *Coding Rights*, 10 September. Available at: <https://medium.com/codingrights/decolonising-ai-a-transfeminist-approach-to-data-and-social-justice-a5e52ac72a96>
- Posada J (2022) Embedded reproduction in platform data work. *Information, Communication & Society* 25(6): 816–834.
- Roberts S (2021) Your AI is a human. In: Mullaney TS, Peters B, Hicks M, et al. (eds) *Your Computer Is on Fire*. Cambridge, MA: MIT Press, pp. 51–70.
- Sinders C (2020) Feminist data set. Available at: <https://carolinesinders.com/wp-content/uploads/2020/05/Feminist-Data-Set-Final-Draft-2020-0526.pdf>
- Suchman L (2008) Feminist STS and the sciences of the artificial. In: Hackett EJ (ed.) *The Handbook of Science and Technology Studies*. 3rd ed. Cambridge, MA: MIT Press, pp. 139–164.
- Suchman LA (1987) *Plans and Situated Actions: The Problem of Human-machine Communication*. Cambridge: Cambridge University Press.
- TallBear K (2013) *Native American DNA: Tribal Belonging and the False Promise of Genetic Science*. Minneapolis, MN: University of Minnesota Press.
- Turkle S (1984) *The Second Self: Computers and the Human Spirit*. New York: Simon & Schuster.
- Wajcman J (1991) *Feminism Confronts Technology*. Cambridge: Polity Press.
- Webb C (2019) WHAT IS A FEMINIST AI? Possible Feminisms, Possible Internets. Disruption Network Lab, Berlin. Available at: <https://www.youtube.com/watch?v=bBQOyvNhWJY>
- Wellner G and Rothman T (2020) Feminist AI: can we expect our AI systems to become feminist? *Philosophy & Technology* 33(2): 191–205.
- West SM (2020) Redistribution and rekognition : a feminist critique of algorithm Fairness. *CATALYST: Feminism, Theory, Technoscience* 6(2): 124.
- Whittaker M (2021) The steep cost of capture. *Interactions* 28(6): 50–55.
- Young J (2019) Why we need to design feminist AI. Available at: <https://www.youtube.com/watch?v=E-O3LaSecVw>

### Author biography

Sophie Toupin is a Horizon postdoctoral fellow at Concordia University where she examines critical perspectives on AI.