Is Artificial Intelligence Gender-Free? What Does Feminist Epistemology Say About That?

Valerija Vendramin
Educational Research Institute, Ljubljana, Slovenia

valerija.vendramin@pei.si

Abstract: I start my contribution with some general questions: Is AI gendered? Is AI sexist or can it be? Does AI include gendered knowledge and suppositions? If so, how? After that, I proceed to develop my theoretical (i.e., qualitatively based) starting points with the main referential authors, Donna Haraway and Alison Adam. The fact that impersonal does not mean observer-independent (as Haraway described it in a slightly different context [1997]) is a good reason to turn to feminist epistemology, especially its concept of situated knowledges. Since knowledge (or the representation of knowledge) lies at the very centre of AI research, this makes it an appropriate vehicle for a gendered critique of AI (Adam, 2000). The concept of situated knowledges entails knowledge that reflects a perspective on a subject which is necessarily partial and limited, not universal (this is Haraway's famous critique of the "view from nowhere"). Namely, there is no way to be simultaneously in all of the epistemologically privileged positions structured by gender, class, nation, etc. I then proceed to AI research. The knowledge engineers build systems that contain knowledge reflecting their own interests and competencies. While this representation of knowledge is usually regarded as being universal (Adam, 2000), it is hierarchical since it does not grant epistemic authority to all. Most importantly, social exclusivism and biological essentialism are re-inscribed in the ontology of AI (Adam, 2000). I address the question of which effects social and political contaminations and prejudices can bring for the development of AI. I suggest that unless we commit to deconstructing the harmful essentialisms that govern our human lives, we might just be perpetuating the same (i.e., our own and others') practices of domination and unequal parts of privilege and oppression (Haraway, 1991) in developing AI.

Keywords: Artificial Intelligence, Epistemic Privileges, Feminist Epistemology, Gender, Situated Knowledges

1. Introduction

As a starting point, I wish to ask some quite general questions about artificial intelligence (AI), a heterogeneous network of technologies, and gender, such as: Is AI gendered? Is AI sexist or can it be? Why? Does AI include gendered knowledge and suppositions? The answers to these questions may seem relatively obvious, at least to those of us engaged in gender research more generally. The notion that AI is completely non-gendered or gender-free is a great misinterpretation, as if AI is being developed in a world devoid of humans, their prejudices (conscious or unconscious), and ways of acting. Since AI is our mirror, why should it be assumed that no sexism or stereotypes will be part of it? Even when AI learns from (existing) data, why should it be assumed that bias will not exist if it is already present in that data? As discussed at the *HeForShe Summit* (UN Women), "AI language models draw their material from already published information, meaning that they can replicate and amplify existing gender biases". In short, what is involved in the ways AI is presented, is a broader social, ethical, and political issue.

I concur with Cave, Dihal, and Dillon's argument (2020, with reference to A. Adam) that it needs to be considered how a certain type of masculinity is inscribed in the way AI is conceived and how this does interplay "with a culture in the computing world that is hostile to women" (their description). AI could hold a more meaningful importance if we agree that AI should benefit all (and also work in that direction). This raises the related question (which I cannot address more deeply at this point) whether this benefit-all goal is reachable under present circumstances. As stated on the webpage of internationalwomesday.com: "Bias may be an unavoidable fact of life, but let's not make it an unavoidable aspect of new technologies". Unfortunately, bias is not something that can be done away with in the field of AI alone, it must be done away with first in wider society. Otherwise, we will just be transferring the same gender-related issues to the new areas (as it is currently done). Nevertheless, the bottom line is: that AI would not contain bias if there were no bias in society. We are therefore confronted with the same old story. Moreover, not only does generative AI replicate stereotypes, it can even make them worse, in many senses, at least in that (once again) an impression is being created that AI is somehow impersonal, society-independent, and objective.

According to Alison Adam (2000), "AI follows classical versions of epistemology in assuming that the identity of the knowing subject is not important. This disguises an implicit hierarchy of knowers involved in the representation of knowledge in AI which privileges the perspective of those who design and build the systems over alternative perspective." Hence, to start dismantling the idea that developing AI is a neutral and gender-free process, as it is too often perceived, we must first return to the very basics of knowledge acquisition and a

feminist rethinking of it. I say feminist because this is where I start from. Yet, I do not suggest that feminist theory was the first or only one to tackle this issue.

2. What About Epistemologically Privileged Positions?

In this section, I present my main theoretical (i.e., qualitative and humanities-based) starting points with the key referential authors, Donna Haraway and Alison Adam. The fact that impersonal does not mean observer-independent (as Haraway described it in a slightly different context (1997), namely, while speaking about reality being collectively, materially, and semiotically constructed), is a good reason to turn to feminist epistemology, especially its concept of situated knowledges. Since knowledge (or the representation of knowledge) lies at the very centre of AI research, this makes it an appropriate vehicle for a gendered critique of AI (Adam 2000).

The concept of situated knowledges entails knowledge that reflects a perspective on a subject, which is necessarily partial and limited, not universal (the famous critique of the "view from nowhere", also known as God's eye view of the world), often concealing a white male perspective (of interest in our case). Traditional epistemology has it as in casting the "business of knowing in terms of 'S knows that p' where 'S'" is universal, and altogether perspective less, taken for granted and not to be discussed subject, and 'p' is a piece of propositional knowledge" (Adam, 2000). This is a view from nowhere, believed to be universal and objective and from everywhere. Alternative (or one of them) is partial, locatable, critical knowledges "sustaining the possibility of webs of connections called solidarity in politics and shared conversations in epistemology" (Haraway 1991). What is furthermore needed is the power-sensitivity and the awareness that there is always the issue of interpretation of results, relations, views, and so on (Haraway 1991).

Here, we find ourselves on the terrain of the necessarily subjective, not objective (objectivity would require a longer elaboration, as there are different definitions of it). Situated knowledges are, in brief, ways of knowing that are "self-reflective concerning the conditions (material, historical, social) under which they came into being" (Prins 1995) and that postulates a critical stance towards 'universal truths'.

There is no way to be simultaneously in all or whole in any of the epistemologically privileged positions structured by gender, class, nation, etc. According to Karen Barad (2007), it is not merely about knowing or seeing from somewhere (as in having a perspective), "but about taking account of how specific prosthetic embodiment of the technologically enhanced visualizing apparatus matters to the practices of knowing".

Although, as said, I do not make any claim that all the topics relating to feminist theory and epistemology fall exclusively within the domain of feminism or feminist theory. They also occur in other epistemologies and philosophical traditions that thrive on scepticism and wariness of absolutes, such as deconstruction, pragmatism, and the like (Felski 2000), I do maintain that it was feminism that bestowed a specific dimension upon them and gave rise to new considerations – particularly to the issues of epistemic responsibility and epistemic privileges. The demand for epistemic responsibility is a central concern of feminist epistemology. "Epistemologists need to take seriously matters of social position, race, gender, sexuality, and the like because social hierarchies can both limit the spheres of action available to agents from non-privileged groups and discourage those from privileged groups from being accountable for their actions when they seek and claim knowledge" (Townley 2006). So this kind of work is necessary to reflect on power, privilege, and accountability, not only in AI research and development, but wider as well.

Or, as Donna Haraway famously expressed (1991), "Vision is always a question of the power to see". This means the assertion made by the researcher (or scientist, author, developer/designer of AI) that they watch from everywhere and see everything, that they have no desires, needs, convictions, or backgrounds, is extremely contentious (Haraway 1991) and an evasion of 'responsible discourse'. This is supposedly a view from a position that is transcendent – as in, above the level of human activity, above politics and power – and beyond lived experience. It is as if these issues have nothing to do with humanity, human practices, and ways of living.

The transformation of recognition systems and methods of observation and, in our case, AI design, requires the destabilization of assumptions concerning the "god-trick" (as Donna Haraway refers to the traditional positivist view of science), where everything is seen from nowhere (from out there somewhere, not from our society), when vision is infinitely mobile and endless (Haraway 1991). As already mentioned, vision is a question of the power to see. A view from a specific location, i.e., from the place where politics and knowledge are being built (Haraway 1991), can enable us to see and understand various aspects of the world and human activities – which are then 'copied' into AI.

For mainstream science thinkers, 'subjective' views are suspect partly because they are associated with embeddedness or situatedness (and sometimes, by extension, with the female). To be embedded is to be immersed in relationships and connections that interfere with the rigor and purity of vision presumably required of empirical science (Greene 1994). As stated by Sh. Pendlebury (2005): "Objectivity requires taking subjectivity into account", which in short means that knowledge is not above the level of human activity with all its values, desires, politics, yearnings, machinations, and so on. Accordingly, "how can a human mired in such a stew produce knowledge that is not" (St. Pierre 2006)?

To put this slightly differently: the stance of the omniscient, universally separated (deemed objective in a different conceptual framework) observer is impossible in our theoretical framework. And that can lead to the conclusion: the observer is always 'somewhere', at a given social location, which simultaneously enables and in a certain way constrains their view (Pendlebury 2005).

This 'obsession' with objectivity and the supposed elimination of bias (through the use of 'positivistic' methods) and overlooking the bias of power must be eliminated from the analysis of AI. There is little recognition of the broader questions of bias relating to which evidence is considered to 'count', and who selects it. Further, the researcher is seen as an outsider working on the social reality, not being a constitutive part of it (Archer 2003) – which can also mean not being responsible for the results and interpretations of data, for example.

3. AI and 'Tacit' Forces at Work

Thus, in my research and reflection on AI, I rely heavily on the work of Allison Adam on this issue as she points out that knowledge engineers build systems that contain knowledge reflecting their interests and competencies. While this representation of knowledge is usually regarded as being universal (Adam 2000), it is hierarchical since it does not grant epistemic authority to all. Social exclusivism and biological essentialism are re-inscribed in the ontology of AI (Adam, 2000). I wish to consider the question of the effects that social and political contaminations and prejudices can bring to the development of AI.

There are several key aspects (as I am told by ChatGPT, which I include here for the sake of illustrating my point (ChatGPT was accessed 14 December 2023, the conversation started with the keywords "gender and AI research"; but, to be clear, this contribution was NOT written by it – I guess such disclaimers will become ever more necessary). It is important, even crucial (says he or she or it), to address gender issues in research and development to ensure that AI technologies are fair, unbiased, and benefit everyone. This proves just the main point. Still, what is fair and beneficial might be a matter of negotiation and depends on the context. The same AI representative (i.e., ChatGPT) also claims that "researchers and developers are increasingly aware of these challenges and are working to develop AI systems that are fair, transparent, and free from gender bias". This is certainly a positive initiative, but – unfortunately, in my opinion – it does not reach very far (or not far enough). An instrumental reflection (I am unsure if I may use the word "reflection" here as this is ChatGPT's words) is the following: "It's important to recognize that any gender-related characteristics in AI reflect the data and methods used in its creation rather than being an inherent property of the technology itself". Yet, how can one thing go without the other? Of course, it is so. There is no inherent property (yet) of the technology. And I am not sure that having an inherent property is a good idea.

To continue (according to the same source): "Gender biases can be a concern if the training data used to develop the model reflects existing societal biases. Developers aim to create models that are as neutral and unbiased as possible, treating all users with respect and avoiding reinforcement of harmful stereotypes".

To comment on the last quote (we clearly would need more data on how many feminist researchers are included in developing the models): "Problematizing the implicit knowing subject of AI through the lens of feminist epistemology reveals elements of the nature of that implied subject which have rarely been exposed in commentaries on AI. They turn the spotlight onto what is known, how it is known, and how that is represented in AI systems" (Adam, 2000). Alison Adam (1995b) also claims – and justly so – that there is "a deletion of the subject in symbolic AI systems. I do not wish to claim that this is deliberate; instead, I see it as largely a consequence of their following in the footsteps of the classical position in epistemology where the nature of the knowing subject has been traditionally denied as an essential element in the making of knowledge". Further, "in deleting the subject these systems thereby delete the political and have no means of making explicit the power structures inherent in the implied hierarchy of knowers". This brings us back to the importance of recognizing epistemological situatedness, which means – precisely – taking subject into account as opposed to the knower in traditional epistemology who is disembodied with reason presented as purely mental (Adam, 1995b). Gender biases are thus a concern and it is very hard to claim that developers are creating models that are unbiased and

avoid harmful stereotypes. Why is that so? Because it is very hard to say that the developers themselves are unbiased (I am not claiming that this is so for some malicious purpose, but still).

In summary: there is bias in AI algorithms (i.e., the training data) which can have a very strong impact if used indiscriminately. There is a thing called algorithmic accountability – which is exactly what it says it is: holding the developers accountable for their decisions (here we return to epistemic responsibility). Certainly, regulations should take care of this, but – once again – we are on the terrain of situated knowledges. A feminist perspective is needed to help develop AI technologies that do not perpetuate the same old issues. Nonetheless, this is again the question not of AI development and technology, but of wider societal issues and the attempt to resolve them (and, in my opinion, we are still a long way from doing so).

4. Conclusion

By way of conclusion, I suggest that unless we commit to deconstructing the harmful essentialisms that govern our everyday human lives, we might just perpetuate, reproduce, and once again legitimize the same (i.e., our own and others') practices of domination and unequal parts of privilege and oppression (Haraway 1991) while developing AI.

As Kate Crawford (2021) states, "artificial intelligence is not an objective, universal, or neutral computational technique that makes determinations without human directions. Its systems are embedded in social, political, cultural, and economic worlds, shaped by humans, institutions, and imperatives that determine what they do and how they do it. They are designed to discriminate, to amplify hierarchies, and to encode narrow classifications". Here, I have not delved into the education, court system, and other specific social contexts that could especially benefit from or be especially harmed by AI. Thorough considerations along those lines will no doubt be needed very soon, if not already (over)due.

AI systems (Crawford 2021) are expressions of power created "to increase profits and centralize control" […] But this is not how the story of artificial intelligence is typically told".

There is no algorithmic exceptionalism, there is no AI that can be more socially just than its human creators.

Acknowledgement

This contribution is a research result of the project, Education at the Frontiers of the Human: The Challenge of New Technologies (EDUCAT(H)UM), financed by the Slovenian Research and Innovation Agency (ARIS): N5-0272.

References

Adam, A. (2000) "Deleting the Subject: A Feminist Reading of Epistemology in Artificial Intelligence", *Minds and Machines*, No. 10, pp 231–253.

Adam, A. (1995a) "Artificial Intelligence and Women's Knowledge. What Can Feminist Epistemologies Tell Us?", *Women's Studies International Forum*, Vol 9, No. 4, pp 407–415.

Adam, A. (1995b) "Embodying Knowledge. A Feminist Critique of Artificial Intelligence". *The European Journal of Women's Studies*, Vol 2, No. 3, pp 355-377.

Archer, L. (2003) "Evidence-Based Practice and Educational Research", in Skelton, Ch. and Francis, B. (eds.) *Boys and Girls in the Primary Classroom*, Open University Press, Maidenhead, pp 26–40.

Barad, K. (2007) *Meeting the Universe Halfway. Quantum Physics and the Entanglement of Matter and Meaning*, Duke University Press, Durham and London.

Cave, S., Dihal, K. and Dillon, S. (eds) (2020) Introduction: Imagining AI, in *AI Narratives. A History of Imaginative Thinking about Intelligent Machines,* Oxford Academic Books, Oxford.

Crawford, K. (2021) *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, New Haven and London.

Felski, R. (2000) "Feminism, Postmodernism, and the Critique of Modernity", in *Doing Time. Feminist Theory and Postmodern Culture,* New York University, New York and London, pp 193–210.

Greene, M. (1994) "Epistemology and Educational Research: The Influence of Recent Approaches to Knowledge", *Review of Research in Education*, Vol 20, No. 1, pp 423–464.

Haraway, D. (1997) *Modest Witness@Second_Millenium.FemaleMan©Meets_OncoMouse™: Feminism and Technoscience,* Routledge, New York and London.

Haraway, D. (1991) *Simians, Cyborgs, and Women. The Reinvention of Nature*, Free Association Books, London.

Pendlebury, Sh. (2005) "Feminism, Epistemology and Education", in Carr, W. (ed.) *The RoutledgeFalmer Reader in Philosophy of Education*. Routledge, London and New York.

Prins, B. (1995) "The Ethics of Hybrid Subjects: Feminist Constructivism According to Donna Haraway", *Science, Technology and Human Values*, Vol 20, No. 3, pp. 352–367.

St. Pierre, E. A. (2006) "Scientifically Based Research in Education: Epistemology and Ethics", *Adult Education Quaterly*, Vol 56, No. 4, pp 239–266.

Townley, C. (2006) "Toward a Revaluation of Ignorance", *Hypatia*, Vol 21, No. 3, pp 37–55.

https://www.internationalwomensday.com/Missions/14458/Gender-and-AI-Addressing-bias-in-artificial-intelligence

https://www.unwomen.org/en/news-stories/feature-story/2023/09/heforshe-summit-discusses-gender-bias-in-ai-and-how-to-encourage-male-feminist-allies