Queering AI

Report 2025

WASP—HS

Table of Contents

Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society (WASP-HS) would like to thank all chairs and participants of the event Queering AI for contributing to the fruitful discussions which this report is based on.

Introduction

In this community reference meeting, we engaged in a transdisciplinary dialogue centered on queer perspectives regarding AI developments, implementations, and discourse. Queer theory provides valuable insights by illuminating often overlooked perspectives, particularly those at the margins, and by exposing the various power imbalances inherent in AI systems. Furthermore, it offers an opportunity to critically examine and challenge the fundamental principles underlying contemporary epistemologies of datafication and automation. By engaging with queer theory, we can foster essential discussions about the potentials and limitations of data-driven technologies across diverse domains such as health, well-being, nature, culture, law, media, and communication.

The background for the community reference meeting was a Swedish network under establishment, emerging through seminar series, mini-conferences, and symposiums, all under the umbrella term "Queer Futures of AI." These events bring together scholars from a variety of disciplines, primarily from the humanities and social sciences, as well as science and technology studies. However, there has been less dialogue with representatives from the surrounding community, including citizens and representatives from the public and private sectors. Therefore, chairing a WASP-HS community reference meeting on Queering AI was a crucial next step to bring these discussions back to society. The aim of the community reference meeting was to highlight often overlooked perspectives and foster critical discussions about the potentials and limitations of AI.

The keynote speaker was Daniella Gati, Lecturer in Games & Interactive Media at the University of Salford in Manchester, UK. Daniella's keynote addressed the epistemology of AI, specifically the shift that AI is causing in how knowledge is understood and created. Daniella discussed the underlying statistical notions of how AI constructs and transmits knowledge and assessed these notions from a queer theoretical perspective. One of the areas discussed in the presentation was dating apps and how generated description texts exert problematic imaginaries and fail to represent the complexity of humans.

The dialogue continued in three parallel round tables

The round table "AI, Data, and Knowledge," chaired by Daniella Gati, discussed the different roles of data and algorithms in how AI creates knowledge and shapes our worldview. During the discussions, participants explored how queer perspectives, both theoretical and lived, could be used to push AI knowledge production in more equitable and just directions. The discussions also related to the keynote, examining the worlds that AI attempts to classify and create, and how these classifications influence individuals' self-perceptions.

The round table "Power and Theory," chaired by Ericka Johnson, explored how we express our identity through our ways of being in the world, how this influences how people categorize us in different contexts, and how this renders identity performative. It was further articulated that these performances are shaped by constraints—power dynamics that define what is possible—and that different contexts and power structures allow for different types of identity performances. Responsibility lays with designers of AI, who must care for the repercussions their AI-design choices might have. Finally, earlier queer movements have resisted prevailed practices of categorization, and we can apply these lessons learned, as inspiration when addressing issues in relation to design and use of AI.

The round table "Large Language Models, Gender, and Us," chaired by Hannah Devinney, addressed the sharp rise in the use of Generative AI over the past year. This rise has sparked conversations about bias, representation, stereotyping, and toxicity. Since these types of AI systems are now applied in a wide variety of contexts, such as chatbots, internet search, virtual assistants, predictive

text, and wholesale text generation, we need to ask ourselves how these technologies will impact the way we experience, talk about, and perform gender. The round table discussed the stereotypes we risk (re-)entrenching through the use of these AI systems and whether there are ways to leverage these tools to mitigate or counter patriarchy and cisnormativity. A pressing concern was who might benefit from these technologies and at whose expense. For participants in the round table discussion that were using AI in practice, there was a mix of excitement and resistance to generative AI, with the risk of stereotyping identified as one significant challenge.

Karin Danielsson and Matilda Tudor,
Community Reference Meeting organizers

Keynote Report Queering AI

Keynote Speaker
Daniella Gáti, Lecturer in Games and Interactive Media, School of Arts, Media and Creative Technologies, University of Salford

I was honored to be invited to deliver the keynote for the WASP-HS Community Reference Meeting on Queering AI. My keynote drew from my ongoing book project that interrogates how AI reshapes the processes through which humans create knowledge. Our knowledge-making has major impact on how we understand our world, our societies, other humans, and ourselves, and as such, on queer ways of being. AI has the potential to change all that—in ways that give some cause for concern.

The presentation examined one particular area in which AI may already be exerting significant influence in reshaping people's imaginaries—that of AI text generation. It used to be the case that much of the text that we encountered on the internet was produced by humans, but nowadays that is changing rapidly as more and more content is generated by chatbots, AI assistants, and the like. That is, these tools have an increasingly large share in creating the environments that allow us to form conceptions of what people are like.

One such area that I discussed in the presentation is dating apps. Before the availability of ChatGPT and such, it was likely that the vast majority of self-descriptions on dating apps were self-authored, or at least written by (other) humans. Thus, a person experimenting with an LGBTQ+ identity, for example, could reasonably assume to learn something about such identities from perusing dating app self-descriptions. But when chatbots make it easy to generate such texts, they contribute to the impression such a person might glean from dating app texts. Now, LGBTQ+ people might all appear to be like what AI text generators "imagine" them to be, rather than reflecting the existing wealth of LGBTQ+ lives and identities.

This influence AI already exerts on our imaginaries is problematic for at least two reasons. First, stylistically, these tools are not able to reflect the kind of diversity that would be desirable for readers to be able to form impressions of humanity as widely varying and colorful. The styles produced by AI may differ to some extent, but they tend to reflect the same underlying model of identity, subjecthood, and language. This is problematic because it reduces the complexity of human experience to a too simplified mold, making it difficult for the human user to gain a deeper sense of diversity. Second and more damning is the particular mold or model into which AI crams human experience. This model is one in which identity is self-evident and frictionless, while language is fully transparent and meaning is communicable without loss. That is, the impression that one gets from perusing AI-generated texts is that the humans who are supposed to speak through them have neither doubts nor insecurities about their identity, nor do they experience flux in them, and they are certain about their knowledge of themselves and their ability to express their straightforward identities directly to other humans.

But this image does not fit well with either how queer theory understands humans, nor with how queer ethics imagines sustainable, responsible human subjecthood and relationships. In queer theory, knowledge, including self-knowledge, is always necessarily limited, and moreover in constant flux: not only are we incapable of fully grasping ourselves at any given time, but even if there was a self to grasp, that self is never static but constantly changing. In line with decades of queer epistemology, I argue that such a model of the human—one that is fragile, fallible, prone to failure and subject to limits—is not only more accurate but also fundamentally necessary for a socially just and environmentally sustainable world. Indeed, many of the crises we experience today—for example that of the climate—owe at least in part to a knowledge-making project that sees complete knowledge as both possible and desirable, no matter the cost. But the belief that humans can and should master and control

everything is clearly misguided, as the acceleration of planetary destruction shows all too clearly.

How, when, and how much AI models are used to generate our world, both textual, visual, and increasingly sonic and video, should therefore be subject to a wide-reaching social discussion, much beyond what is happening at present. The model of knowledge that AI perpetuates and extends is one in which humans can and should seek full mastery over knowledge, language, identity, and world. Indeed, it is this very form of knowledge, one whose end goal is full mastery, that underpins the project of AI development itself—a form of knowledge that does not pause to ask questions about responsibility, failure, limits, and loss. Not coincidentally, not only the training but also the use of AI models amounts to an irresponsible consumption of water and energy—another aspect of how the project for domination sidesteps the question of limits, failures, responsibility, sustainability, and justice.

A queer vision for AI would entail embracing limits, failure, and loss. It would mean sacrificing some accuracy in order to allow more of the unpredictable, while letting go of accuracy goals that lead us toward planetary destruction. It would necessitate accepting that total control cannot be our final aim, and that machine-produced language can never communicate human feelings, thoughts, and identities—not even humans can.

Roundtable Discussion

Large Language Models, Gender, and Us

Author

Hannah Devinney, Postdoctoral Researcher, Department of Thematic Studies (TEMA), Linköping University

Main Challenges

- Large Language Models (LLMs) and other generative AI systems can be used at large scales, meaning many people will have repeated conversations with them, and they are often presented with the illusion of personhood. This gives them power to shape and reify the ways we talk and think about the world.
- Knowledge distribution about such systems is uneven, and even with a desire to use them "in a good way" participants are unsure of how to connect all the pieces in order to bridge the gaps.

Generative AI in the language domain, such as ChatGPT, have recently experienced a sharp rise in use and popularity, opening conversations about—among other concerns—bias, representation, and stereotyping. These systems are now being applied in a wide variety of contexts, including chatbots, text generation, and virtual assistants. The purpose of this roundtable was to explore the impacts of these technologies on how we experience, talk about, and interact with identity categories such as gender. Our discussion ranged from our fears to our hopes for the uses of these technologies, often revealing tension within and between these wants and concerns.

Paths Towards Possible Futures

For those who work in areas where the incorporation of generative AI is being encouraged, there was a sense of being caught somewhere between "two sides." Participants felt that some in their organizations were really excited about these technologies, while others were very afraid or resistant to using generative AI at all. A stage of wondering—how to use Large Language Models well? What is valuable to us? How will this change how we work? —was articulated, along with concern about knowledge gaps and how expertise was regarded and, consequently, responsibility assigned.

For example, one participant mentioned that a shifting of responsibility to "techy experts" and out of the hands of those who are used to doing the work seemed to produce a gap within the organization, and a lack of knowledge about bias on both sides: one missing the technical knowledge of how biases occur in the algorithmic systems, the other missing the contexts around inequalities in the population the system is meant to serve.

Inequality, Diversity, and Stereotyping

Inequality and diversity were the red threads of the discussion, with Large Language Models positioned as potential drivers of both. On one hand, they could potentially be used to create synthetic diversity, such as fictional "test sets" for new survey questions or probing other systems for bias. On the other hand, there are clear concerns about reifying bias, and there is an obvious tendency for Large Language Models to stereotype and only produce superficially diverse content.

Even when diverse options are available, the default representations of gender in systems such as chatbots or virtual assistants tend to be very stereotypical. The textual domain was described as particularly challenging because in this modality it is only the words that "give away" identity or personality. This may mean that it is easier to miss subtle, underlying stereotypes. Defaults are important because they are "chosen without choosing". Combined with the scale Large Language Models can operate at, the sheer amount

of exposure and repetition can work to reinforce gendered stereotypes and other biases.

What Gives Large Language Models Power?

Automated systems are in general lent power by their scale. Although one short encounter with a Large Language Model may not be experienced as biased or stereotypical, repeated conversations can reveal issues with patterns of behavior, such as the "scripting" of identity. Anthropomorphizing also lends a legitimacy to systems such as chatbots: when we think of one as a person it becomes part of the co-construction of our understanding of the social world, and chatbots that "sound like" people are often taken as thinking entities. When these entities are encountered and play into stereotypes (such as presenting a friendly, feminine persona as a personal assistant), they reinforce inequalities.

Continuing Questions

We left the roundtable having explored different concerns and possibilities of Large Language Models, which prompt questions about their use. How can we make the best use of these technologies? How does one find what is good? Who should be trusted, or given responsibility as "the expert"? When the system feels so amazing as-is, how can we keep the layers behind it clearly in our minds while making decisions?

As one participant expressed, the feeling in their team was "we should use [generative AI], but with caution." These technologies represent strong potentials for contributing to the more diverse, more equal futures we want to create. However, the risks are also great, and cannot be ignored if we want to use Large Language Models and other generative tools in a good way.

Roundtable Discussion

Power and Theory

Author

Ericka Johnson, Professor at the Department of Thematic Studies (TEMA), Linköping University

Main Points:

- There is a conundrum between wanting guardrails to protect safe spaces and, at the same time, creating spaces that allow for subversive and queer potentials.
- The designers of AI have the responsibility to think about the repercussions of their technology-design choices.
- Queer movements have resisted privileged practices of categorization before. We should look to past activism campaigns for lessons-learned and inspiration.

This conversation started with the reflection that Queer theory encourages us to think of identity as performative and power as constitutive, ideas which challenge a paradigm that imagines data as representative of a person as stable and separate from structural power dynamics. Beginning there, we collectively mused about what queer guardrails would be needed to allow for performative identities. And how could we use reflections on power rather than identity to shape these guardrails?

AI Guardrails

These starting points brought us into many different lines of thought. One, of course, was about how AI guardrails (and AI tools in general) could be used to protect against abuse and harmful behavior. This doesn't even have to be intentionally harmful behavior; it can be just the space and technology to reproduce (dominant) interaction patterns that privilege some voices over others. Surely there would be guardrails which could prevent or mitigate this.

However, as much as we recognized the need to create safe spaces that protect people from abuse and allow for equitable participation, we were also wary to employ guardrails or technologies that may limit and filter away subversive and queer potentials. This conundrum — between guardrails and the flexibility to engage with tech in one's own, desired and desiring manner — was exemplified with the chatbot Replika, which is well-trained to not cross boundaries or become abusive, but which was revamped with stricter guardrails around the time of Valentine's Day 2023, to the dismay of many users.

The Responsibility of Designers of AI Tools

Another recurring point in our conversations was that the designers of AI tools have a responsibility to their users and for the technology they design and develop. This reflection encouraged us to discuss different ways design can enable efficient use or, in a more guardrail-sort of spirit, design models that engage friction and make the AI experience less convenient, less efficient. Here there is a tradeoff between 'real' and 'inconvenient', and where AI should end up in that balance is not a given for every context. This responsibility could even be stretched to include consideration of sustainability issues, including the energy consumption of AI.

Finally, we noted that there have been many times in the past when activist strategies were employed to challenge, among other things, categorization processes and the relationship between technology and the reproduction of privileged people's points of view. We should look to these movements for lessons learned and inspiration.

Roundtable Discussion

AI, Data and Knowledge

Author

Daniela Gati, Lecturer in Games & Interactive Media, School of Arts, Media and Creativ Technologies, University of Salford

Data Categorisation

The roundtable was comprised of a varied range of participants with different backgrounds and research interests. Participants were invited to discuss how AI creates knowledge and shapes the world as well as how queer perspectives both theoretical and lived can be used to push AI knowledge production in more equitable and just directions.

The first topic discussed was data categorisation and how categorisation runs the risk of stereotyping and generalising. Data categorisation is necessary for many tasks involving large sets of information, whether this is computational analysis of texts or images or even such simple tasks as finding particular items or topics in a library. Data categorisation enables the organisation of and access to information. But categorisation invariably produces a particular vision of the world in which the categories appear as intrinsically distinct from one another. The question of how to mitigate these risks was raised and discussed. One participant shared how an effort to catalogue Swedish LGBTQ+ literature (Queer Lit) tried to navigate these risks, while another pointed out that categories often lead to LGBTQ+ content being earmarked as "bad" content and therefore erased out of view. One participant further argued that responses to these risks from industry and the developers tend to deflect away from any responsibility their AI tools might have: on the one hand, they often claim that AI has all the solutions, but when it comes up short the response is that it is unfair to expect anything more from AI than the flawed responses it is already capable of delivering—in other words, that it is the expectation of a more nuanced and equitable categorisation from AI that is itself unfair.

The discussion then turned to a related issue with categorisation/classification, one that was closely linked to Daniella Gáti's keynote: this was the issue of how the attempt to classify not only describes a pre-existing world, but in fact also produces the world that it describes, and in so doing flattens the world. Representation influences how people see themselves, and especially when representations tend to be alike, it becomes difficult to conceive of oneself in varied ways. This is a problem that participants discussed in terms of the difficulties with making LGBTQ+ people and concepts more visible. As an example, Homosaurus.org was mentioned, an online, international LGBTQ+ linked data vocabulary, which responds to the need for and importance of increasing the visibility of marginalised people by creating a catalogue of LGBTQ+ vocabulary. However, Homosaurus.org exemplifies the challenges in well-intentioned and important awareness-raising because its kind of categorisation also constitutes a limitation—a particular form of knowledge, one of many ways of describing the world which will in turn shape the world. Another example that was mentioned was how to represent queerness on Wikipedia.

Both of these efforts run up against the problem that information representation (via categorisation, classification, metadata etc.) is never just representation, but also information creation. Queer theory has long addressed the injustices with which claims of describing preexisting knowledges actually produce particular ways of knowing, and the produced knowledges enforce particular ways of being and police others. Judith Butler's famous point was mentioned as an example: Butler points out that the social world has been divided according to reproductive function, an odd and random form of classification that then produces a knowledge (about social roles and responsibilities) that is taken for granted as "natural," as simply a fact.

It is not only the particular forms of classification that can be challenged, but, as queer theory does, classification itself. Such challenges are exempli-

fied by projects to represent people who resist classification, such as the Wikidata project referenced above. There is a paradox here, where classification is necessary and avoiding classification is also necessary at the same time. It remains an important question whether this problem can ever be solved, or whether all that we have at our disposal are ways to mitigate it.

Current Frameworks

The discussion then turned to our current frameworks for addressing this problem. Currently, much of computer science and industry simply proposes "more tech" and "more data" as the solutions to every problem. Participants were instead interested in exploring limits and vulnerability. Butler's work on precarity as a founding condition of human and non-human life suggests that accepting the limits to our projects of knowledge may be a more equitable form for technological development than the currently predominant visions.

The current socially dominant imaginary, spearheaded by the companies profiting from AI and uncritically echoed by many governmental and nongovernmental institutions, is one of AI as a boundless technology with no limitations on its (current or future) ability. This vision of AI underpins the techno-utopian belief, articulated as a conviction, that AI can do anything and solve everything, and that therefore it is "more AI" that must be our answer to every problem. This belief system or ideology reinforces the notion that placing limits upon AI in any way is fundamentally wrong, and it manifests in projects such as AI's limitless harvesting of our data. In the face of the assumed need for unbounded AI expansion, questions of individual privacy, intellectual property, consent, and equitable use of data become irrelevant. How, how much, and when data is collected all remain undisclosed, and why that may be problematic is not even discussed in the face of the presumed primacy of AI expansion over everything. Limits are absent.

But participants pointed out that such an expansionist, techno-utopian, and accelerationist view is flawed on many accounts, primary among which perhaps is its misrecognition of limits. Limits, as Amanda Lagerkvist argues in her book Existential Media: A Media Theory of the Limit Situation, are a fundamental part of our existential condition: we cannot transcend them, and their existence yields meaning. Furthermore, limits unite humans and non-human living beings in a shared condition of precarity, as Butler argues in their work. Overcoming human limits, or the limits of life itself, are thus not desirable—moreover, such projects share an alliance, if not an open one, with fascist dreams of improving the race.

Additionally, as participants pointed out, our planet is itself not without limits. Notably, the role of AI in humanity's averting of those limits or accelerating towards them is one in which participants once again pitted themselves against a mainstream vision. For the latter, AI's role for the planet, if addressed at all, is again one of unbridled potential: AI will, so the promise goes, be the thing that solves the unsolvable puzzle of the climate crisis. What remains unacknowledged, though, as participants pointed out, is the extreme extent of the planetary costs AI is inflicting already now—both in model training and, controversially, in its use by individuals. Notably, the generation of 100 words by current models consumes just over 500ml of water, and the equivalent of 14 LED bulbs burning constantly for an hour. So while perhaps AI may be used in research to develop applications that can mitigate climate damage, its current frivolous use—which is encouraged if not mandated by the large tech companies—risks destroying the planet before we get a chance to develop climate solutions. Once again, limits are painfully absent here.

In this regard, participants expressed concern not only about AI's climate impact but especially about the general lack of awareness/initiative about in in the public, in industry, and in research. Connections to academic research were raised: research is problematic when it is seen as an end in itself, where "more" knowledge is regarded as always good and therefore justifies limitless resource use. Researchers can place limitations on their own work, as exemplified by Daniella Gáti in the keynote; but ultimately, the public and industry must also place limits on their use and development of AI.

The Need for Limits

If it became increasingly clear during the roundtable discussion that limits are essential for jus-

tice, flourishing, and indeed survival, the question of how to introduce limits and uncertainty remained difficult, with no easy answers. Participants discussed how, in the few instances when limits are imposed in tech, these tend to be misapplied and exclusionary. For example, filters intended to safeguard against bias often tend to diminish nuance and erase LGBTQ+ language, with the result that only the normative is left. Indeed, within tech there is often a deference towards certain discursive projects such as the retrograde side of the cultural wars. Such instances exemplify limits that are misapplied, operating by the rule of large numbers and thus keeping queer, black, differently abled and non-normative people on the margins.

Thus, ultimately participants were left with an acute diagnosis of our current paradoxical moment. On the one hand, following from the techno-utopian idea of AI being able to do anything, AI is heralded as the solution to everything, including the climate crisis, and AI acceleration is therefore seen as necessary and to be prioritised ahead of all other issues, including those of representation, civil society, human rights, etc. This view regards any obstacles to acceleration as contributing to the crisis and leading to destruction. On the other hand, this view, which is the currently dominant one, fails to acknowledge the existence of human and planetary limits, and sets the wrong priorities, regarding business incentives as more important than the survival and flourishing of all people and the planet. As one participant noted, this situation presents us with dead ends everywhere we turn, since AI's acceleration not only exacerbates inequality and oppression but also accelerates us towards the ultimate planetary crisis and our own self-destruction.

At the end of the roundtable, it was asked if a fully equitable, liberatory, just AI is possible or ever will be possible at all. Participants were sceptical. Regardless of who uses it, bias will always be present; categories are not easily overcome. Moreover, it is clear that AI is currently not working for marginalised people, average people, nor the planet, while it accrues power and profit to those who already have a larger share in them. However, participants still ended on the note that despite the risks, scepticism, and difficulties, AI should not (perhaps cannot) be abandoned. Even if we don't believe that the technology can ever be fully fair, we can, and indeed must, work towards making it fairer and better, while accepting and acting on our responsibility toward the survival of the planet. Working with limits may be a way forward.

References

Beatrice Melis, Chiara Paolini, Marta Fioravanti, Daniele Metilli, What Does It Mean to Be Queer in Wikidata? Practices of Gender Representation Within a Transnational Online Community, Communication, Culture and Critique, Volume 17, Issue 3, September 2024, Pages 200–207 https://doi.org/10.1093/ccc/tcae029

Chen, S. The Lost Data: How AI Systems Censor LGBTQ+ Content in the Name of Safety. Nat Comput Sci 4, 629–632 (2024). https://doi.org/10.1038/s43588-024-00695-4

Verma, P., & Tan, S. (2024, September 18). A Bottle of Water per Email: The Hidden Environmental Costs of Using AI Chatbots. Washington Post. https://www.washingtonpost.com/technology/2024/09/18/energy-ai-use-electricity-water-data-centers/

This Report

This report is made possible by the Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society (WASP-HS), a national research program in Sweden. The vision behind WASP-HS is to promote new interdisciplinary knowledge in the humanities and social sciences on the subject of artificial intelligence and autonomous systems and their impact on human and social development.

For more information please visit www.wasp-hs.org.

For questions or inquiries please contact us at contact@wasp-hs.org.

Author

Daniella Gáti, Lecturer in Games and Interactive Media, School of Arts, Media and Creative Technologies, University of Salford

Ericka Johnson, Professor at the Department of Thematic Studies (TEMA), Linköping University

Hannah Devinney, Postdoctoral Researcher, Department of Thematic Studies (TEMA), Linköping University

How to cite this report

WASP-HS. Queering AI. Report 2025.

WASP—HS

The Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society