# AI TO PROMOTE GENDER EQUALITY

Adressing gender inequality through AI – a learning summary

VINNOVA
Sweden's Innovation Agency

Authors; PhD Sophia Ivarsson

# INTRODUCTION

In the development of AI, several studies have shown that bias (prejudice / bias / distortion) in general and gender bias specifically, are present in all stages of the development of AI and arise due to several different factors, such as the lack of diversity in development teams, limited and/or biased data, the design of algorithms and decisions related to how data is collected, coded, or used to train algorithms. One effect of this is that AI-related services can provide up to 70% lower quality to women and non-binary people, according to analyses performed at the Center of equity, gender, and leadership (EGAL) at the UC Berkeley Haas School of Business.

As the discriminatory effects of AI applications are increasingly discovered and trust in AI decreases, research in the field of AI and gender bias accelerates. The results from these show that the gender perspective is rarely addressed specifically in the development of an AI solution, but only included in other routines / processes, usually related to data protection.

Being able to understand and explain the reason for inequality in AI solutions is especially important as the cause of inequality in some cases can be an effect of prevailing inequality in society and thus can be said to be "real" or "reality-based". But they can also be process related. Thus, gender analyzes need to be included in AI development from the outset and in all parts of the development process. To get rid of the problem of gender bias in AI development, by removing gender as a variable has been encumbered with several problems. Not least because other variables can correlate with gender and thus give the same biased outcome, only more difficult to detect. Another significant problem is that the potential positive effects on gender equality aspects are lost.

Instead of focusing on the gender bias problem in the general AI development, this strategic program, AI to promote gender equality, aims to highlight the positive effects on gender equality that AI can bring, by examining how AI can be used to address gender inequality challenges.

Since 2019, Vinnova has been developing the strategic program "AI to promote gender equality". Its purpose is to investigate how AI can be used to help solve gender equality challenges. By focusing on how AI could be used as a tool to solve gender equality challenges, negative effects of AI technology, linked to gender bias can be solved in new ways, for example by highlighting the contextual contexts that require deeper gender analyzes of data but also access to other types of data than those made available in other types of AI applications.

The ambition is also to bring AI into a field, the field of gender research, that is not usually in the forefront of experimenting with and developing emerging technologies. A field that is also very female dominated.

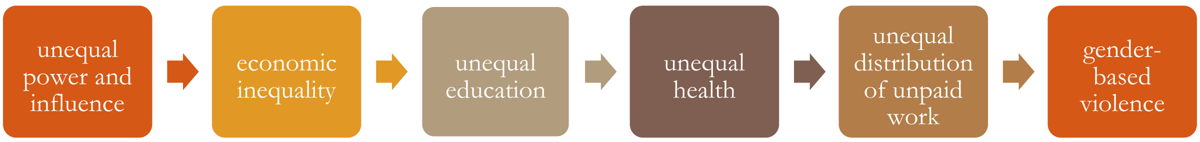
From a gender perspective, it is important that the development of new technology covers areas where women are largely active. Historically, Tech areas have had a very skewed gender distribution and have often been developed in male-dominated areas, something that risks being repeated for each new technology area; thus, more formative efforts are needed, rather than "correcting" problems afterwards. With this program, we hope to be in the forefront and bring new knowledge to the field of AI as well as the field of gender. And to connect and mobilize experts in both fields, because this will certainly be needed in the future. As much as gender expertise can bring important knowledge to the field of AI, AI can bring new, important knowledge to the gender field.

This report sums the results obtained from our project. It is based on the activities performed in the project; a call focusing on how the SDGs focusing on equality can bring innovative solutions with new technology; two international hackathons bringing actors from different fields of expertise together to solve gender inequality challenges with AI and workshops with Women in AI focusing on national gender inequality challenges and how they could be addressed through AI.

A learning project was set up in parallel to the activities to capture results and interviews with participants from the activities was performed. This report presents results from the interviews performed during April 2022 to May 2022.

## Identifying gender inequality challenges|

The gender inequality challenges that was used as a framework (see below) are the Swedish gender equality political goals. However, they also correspond to gender inequality challenges identified by EU and UN. As such, they are of global relevance.  In our activities, participants were instructed to develop AI solutions related to these gender inequality challenges.




## Interviews|

A survey to the participants was distributed, both those who had received funding in our calls and participants from our hackathons. The projects that were identified from the survey as the most developed were interviewed with in dept questions. In all, these were 7 projects focusing on:

1. AI as a tool to mitigate bias in venture capitalists' decision making.
2. AI as a tool to assist a gender equal recruitment, salary setting and promotion.
3. AI as a tool to obtain gender equal pension.
4. AI as a tool to mitigate gender discrimination and stereotypes in children's books.
5. AI as a tool for inclusive and gender sensitive communication.
6. AI as a tool for identifying drivers and obstacles for inclusive workplaces.
7. AI as a tool to monitor speech time, interruption, and verbal discrimination in workplaces.

Other interesting projects such as an AI tool to improve cardiac diagnostics for women, an AI tool to detect and suggest equal division of domestic chores, AI to assist early neuropsychiatric diagnoses in young girls, are also results of this project however, they were in earlier phases of their AI development at the time of the interviews. In the second hackathon, an intersectional, normcritical approach was added to the framework. Evaluations of this expanded focus remain to be performed.

**Combining Gender & AI = A hit!**

It is commonly assumed that the problem with gender bias in AI development is due to lack of diversity in project teams, and the underrepresentation of women in AI is well known. Whereas this project does not focus on the underrepresentation of women in AI per se, it was a hypothesis that it matters for women what you focus on when developing or applying Tech. This was also supported by the respondents, as they said it was the combination of the focus on gender equality issues, combined with AI that evoked their interest and decisions to participate in the project.

> *"I´m curious about AI but don´t know so much. Gender equality, though, I know something here and I have knowledge to bring into this area. That was both challenging and safe."*

Several of the participants had no previous knowledge in AI but were engaged in gender equality issues. It was emphasized by some of the respondents that it is not necessary to be an AI expert to engage in finding solutions to gender inequality challenges. You need to understand some basics but in the first phases you can experiment with data according to your ideas. The threshold is lower than many people might think when AI is involved.

On the other hand, participants with no previous knowledge in gender, testified that they had gained important insights in gender inequality issues. Whether one came from the field of AI or gender, the teams had gained knew knowledge which strengthened their expertise.

**The lack of data problem**

A common challenge for the respondents was lack of relevant data.

The lack of data problem is found on several levels.

1. Access to data - Since many gender inequality challenges focus on individuals there is a challenge with integrity issues and an unwillingness by companies to provide data. To only have access to open data limits the possibilities, hence real data from companies is needed.

2. The digital gender divide – the use of technologies that generate data about their users are more common among men, thus, there are fewer digital footprints from women, which might skew datasets. Further, the collection of data might be skewed, in medical testing, for example, women have historically been excluded in medical testing, a gap that will be reflected in medical data.

3. Disaggregated data – lack of sex disaggregated is yet another problem as it paints an inaccurate picture, concealing important differences between women and men and other categories, and thus hides potential overrepresentation or underrepresentation in the data set.

4. Regulatory constraints - gathering and analyzing valid data poses ethical questions and privacy concerns. This is specifically prevalent in AI applications promoting gender equality. To obtain gender disaggregated, personal data that, at the same time, preserves privacy, poses challenges. The use of synthetic data has been suggested as a remedy. However, some research show that synthetic dataset that are close to the original data set tend to reveal the original data subjects. The use of synthetic data generation to add or remove gender data might not be sufficient to address gender bias, since proxies that correlate with gender will still present problems.

5. Pre-annotated data - when using pre-annotated data there are a lot of gender biases which makes them difficult or impossible to use. It is also problematic that much data has both second and third levels of gender connotations, which means that gender biases occur even though gender as a variable is removed.

*“-a blue bike was connotated with masculinity, as well as blue sheet. Blue seems to be more strongly connotated with masculinity than the objects bike and sheet. Bike and sheets alone, though, would have had other gendered connotations.”*

As illustrated above, to use pre-annotated data, gender connotations must be thoroughly analyzed before data can be applied, which is very time consuming. Yet another problem that was raised with large databases was the lack of gender as a factor in these data sets. If you want to bring in gender classified data, as would be the case when working with AI solutions to promote gender equality, it becomes difficult if the data set do not have gender classifications.

**The importance of gender analyses.**

Having humans in the loop of developing AI solution is of course even more important for AI solutions with the ambition to mitigate gender biases. To work with AI solutions that promotes gender equality it was also emphasized by the respondents that gender analyses had to be performed by gender experts, especially when it comes to choosing data, which needs to be done based on gender research. It is also important to *understand* data from a gender perspective and to validate the algorithms on how it performs on gender relevant outcomes.

> *“At least twenty of the algorithms most influential variables should be verified as to their effects on gender relevant outcomes.”*

Moreover, analyses on gender relevant outcomes needs to be done as soon as any new variable is introduced. Dataset today are commonly tagged as either one or the other gender, but words are extremely context dependent, and getting around that was very challenging. As a hypothesis, the gender of an author of a text, was believed to determine how the text would be classified by gender, however it turned out that the type of language, for example, academic texts, were more influential in how the text was classified by gender. Data verified by linguistics knowledgeable in gender aspects would have made a big difference for several of the participants.

**Resistance**

As with many gender equality investments, resistance come in to play. We therefor asked the respondents if they had experienced resistance in their projects. Of those that testified to such experiences, it was described as resistance from men that could lose some of their privileges if the AI tools were to be implemented. An unwillingness, by older men, to invest in AI tools to solve gender inequality issues was also experienced.

> *We competed in a startup competition. We are all PhD students in Tech, yet they scored our technical competence 2/5.*

**When the hurdle becomes the solution**

The hurdle with gender bias can in fact be the big advantage. Many of the respondents describe that the benefit with gender biases when it comes to applying AI as a tool to address gender inequality challenges. A whole new universe of gender biases, previously not known to us, can be revealed in data, and used to mitigate gender biases that causes discrimination. AI makes it possible to see which of the variables are the most important in affecting gender outcomes. However, interpreting these gender outcomes requires experts in the field of gender studies.

The project developing an algorithm to assist gender equal decisions in startup investments, validated the AI tool that corrected for unconscious gender bias. It was found that the AI tool rendered a higher economic rational, i.e., higher returns of investments. Hence, it is financially smart to correct for gender bias in VC investment decisions.

> *if the startup was a high technical solution and run by an attractive woman, it resulted in lower VC investments. The opposite was found when the startup was led by a man.*

# Summary|

The project AI to promote gender equality has created new insights both within the field of gender as well as the field of AI.

First, we see promising evidence that AI and automation can be a scalable tool to promote gender equality and overcome discrimination. As AI models produce artifacts based on access to objects, categories, properties, and relations between them, they comprise, represent, and reinforce a broad range of different ideas and values, power relations, stereotypes, and prejudices. With AI, these relations can illuminate interconnections previously unknown to us. Used in this way, AI offers unique insights about societal gender regimes, as well as a tool to mitigate these.

This, however, requires expert knowledge in terms of validating and interpreting these relations from a gender perspective.

Challenges to the development of AI solutions that address gender biases, is the lack of data that allow for models to be trained and developed on. Access to relevant data requires obtaining permissions to private data that are subject to different confidentially rules. Moreover, real world data often contains gender biases, because of inequalities in current societal structures, and building models on biased data carry the risk of reinforcing these very biases. Finally, organizations may feel reluctant to share data that may exposes systemic problems in the way they operate. These are issues that all the projects that Vinnova has financed in the field of AI for gender equality have encountered. However, a relatively new development in AI may bring some solutions to the scarcity and inadequacy of data to develop gender balanced models.

Synthetically generated data consists of completely new and artificial data points with no one-toone relationships to the original data, but synthetic data points cannot be traced back or reversed to the original data. There are several ways to create synthetic data. One consists of a mixture of data from different people. Another way is to put noise on existing data, to avoid tracking the data. A hybrid synthetic data can also be obtained by first creating synthetic data and then points in the original data being most similar to the synthetic data points, create a mixture of the original data and the synthetic data. With synthetic data domain knowledge can be injected into the training of AI models to improve the quality of the model's prediction.

Synthetic data can also train AI for new scenarios not seen before, which is especially interesting in the field of gender studies as gender equality, in many domains, is yet to be attained.

Thus, synthetic data carries a lot of promises. According to Gartner, 60% of all data used in the development of AI will be synthetic rather than real by 2024. It has been used in healthcare where synthetic data -from artificially generated medical histories to healthcare claims - have been generated in order to develop, improve and scale AI use cases. Synthetic data preserve the structure and "signal" of real patient data while directly addressing the data privacy concerns that for years have held back the deployment of AI in healthcare.

Applied to gender inequality challenges, synthetic data can be seen as a secure substitute to real world data in that it respects personal data protection regulations and non-discrimination legislation. This could ease down some of the barriers that prevent companies to share their data with researchers developing AI solutions to address gender biases. However, special attention needs to be given to how synthetic data is generated. Indeed, if synthetic data is a mirror of the real world, and real data world is biased, then synthetic data will reproduce this bias. Fixing imbalances and bias are complex matters. Even if a dataset does not contain data on sex, gender or other information related to protected grounds, such as ethnicity, sexual orientation etc., bias and discrimination might still occur due to proxies that correlate with these protected grounds.

A newly developed field called fair synthetic data offers the possibility to fill out data set with potentially more representative data and claim that bias can be corrected. Researchers at Amazon published a report in 2021 providing evidence of a method that produce fair and unbiased data and used it to train AI models. However, there are limitations also to fair synthetic data, for instance if there is a lack of representative research on gender differences in medical diseases, synthetic data couldn't correct for that. In addition, these methods require high computation resources. How synthetic data could be useful in AI applications that promote gender equality need to be explored in more detail.

With the rapid developments of AI there is an urgent need to involve domain experts with profound knowledge of the social, cultural, political and economic artefacts produced in data, that can be readily activated in different use cases.

Innovative solutions on new forms of collaboration are now being funded by Vinnova and coordinated by AI Sweden. The aim is to create a platform for knowledge exchange between problem-owners, AI-developers, and domain experts. Researchers from social sciences and humanities as well as civil society organizations will form a cross-disciplinary expert pool and explore how domain experts can collaborate to support more nuanced decisions and accurate solutions that increase usability, fairness and trust in AI models.

As a result of this learning summary, when the strengths of human and artificial intelligence are combined in a hybrid intelligence approach, it can indeed invigorate the journey towards gender equality.

# RECOMMENDATIONS

- **A multi-stakeholder, transdisciplinary approach is necessary and gender experts need to be involved from start in the development process**
- **Choices and collection of data must be based on gender research.**
- **Gender and intersectional analyses of AI systems need to be systematically performed in pre-, and post-processing phases and as soon as new data is introduced.**
- **Identify, verify, and communicate at least 20 of the most influential variables driving the algorithm on how they affect gender outcomes.**
- **How synthetic data can increase opportunities to develop AI that can promote gender equality should be explored.**
- **Research and innovation funding agencies (RFO) should require applicants with AI proposals to perform gender and intersectional analyses and integrate these as part of excellence criteria in the evaluation process.**

## AI METHODS AND TECHNIQUES USED BY THE PROJECTS |

1. AI as a tool to mitigate bias in venture capitalists' decision making.
    a. The project trained a machine learning model that predicts the return on investment of investments in high tech companies based on a dataset from an investment platform. The model offers predictions that are based on economic indicators, and therefore mitigate for gender biased that may interfere with the evaluation of potential investments by investors.
2. AI as a tool to assist a gender equal recruitment, salary setting and promotion.
    a. The project reveals spreads of salary for men and women given a range of indicators (seniority, education, experience, responsibility, time off due to sick leave or parental leave. Etc.) and offers a factual basis for salary negotiations.
3. AI as a tool to obtain gender equal pension.
    a. The project is a Fin Tech app to help couples save together to reduce the pension gap between women and men. The plan is to use Robo management to identify

and highlight new ways of equal savings with information that the banks don´t have, for example differences in spendings on private vs family expenditure.

4. AI as a tool to mitigate gender discrimination and stereotypes in children's books.
    a. The project has used NLP analysis to identify which words that were mostly associated with men and women. The results of the analysis are displayed in a web app which visualizes the gender biases in word clouds. The results are available for publishers to make informed decisions.
5. AI as a tool for inclusive and gender just communication.
    a. NLP analysis was used to identify which words and expressions were more associated to men and women. 2 key datasets were used as the source for the analysis: TAKOM-1 (the project own database constructed from data extracted from Blogs targeted at Men and women) and PANDORA (database of reddit comments). Text was categorized for biases at words/phrases/paragraph levels using a variety of machine learning models. The models were then applied on new text, and those text analysed for biases. The biased words were replaced by similar words, and the text rescanned for biases until a satisfactory phrase free of biases was obtained.
6. AI as a tool for identifying drivers and obstacles for inclusive workplaces.
    a. The project looked at communication patterns and used social network analysis to identify the power dynamics in organisational groups in conjunction with statistics on sick leave to reveal gender differences and unhealthy work patterns.
7. AI as a tool to monitor speech time, interruption, and verbal discrimination in workplaces
    a. The project applies sound recognition techniques to identify speaking time by different participants in a meeting and gives real time feedback to the meeting participants.

# REFERENCES

Antretter, T., Blohm, I., Siren, C., Grichnik, D., Malmström, M. & Wincent, J. (2020). Do Algorithms Make Better - and Fairer - Investments Than Angel Investors? Harvard Business Review

Bent, Jason R., Is Algorithmic Affirmative Action Legal? (April 16, 2019). Georgetown Law Journal, Forthcoming, Stetson University College of Law Research Paper No. 2019-6, Available at SSRN: https://ssrn.com/abstract=3372690

Gendered Innovations 2: How Inclusive Analysis Contributes to Research and Innovation (europa.eu)

Johansson, J., Malmström, M., Lahti, T. & Wincent, J. (2021). Oh, it's complex to see women here, isn't it and this seems to take all my attention! A repertory grid approach to capture venture capitalists' cognitive structures when evaluating women entrepreneurs. Journal of Business Venturing Insights, 15, Article ID e00218.
(PDF) Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning (researchgate.net)
Smith, G & Rustagi, I. 2020. UCB_Playbook_R10_V2_spreads2.pdf (berkeley.edu)
When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity (ssir.org)
White, Andrew (2021). By 2024, 60% of the data used for the development of AI and analytics projects will be synthetically generated. Gartner Blog Network.
Weijie Xu, Jinjin Zhao, Francis Iannacci, Bo Wang (2021) ffpdg-fast-fair-and-private-datageneration.pdf (amazon.science)