

## **Gender Bias in Voice Assistants**

*Reproduction of Stereotypes through Artificial Intelligence*

Michelle Kramny, Bergische Universität Wuppertal '27

**Abstract:** This paper investigates the reproduction of gender bias in voice assistants, a prevalent application of artificial intelligence (AI) increasingly integrated into everyday life. Despite their apparent neutrality, voice assistants often embody and reinforce traditional gender stereotypes, primarily through the widespread use of female voices that exhibit submissive and service-oriented behaviors. Drawing on feminist philosophy of science, this study critiques the assumption of objective and value-free technology by highlighting how socio-cultural power structures influence AI development. Key factors such as biased training data, algorithmic design, and the male-dominated technology sector contribute to the perpetuation of gendered norms in AI systems. Through an analysis of linguistic interaction patterns and technical foundations, the paper demonstrates how these biases manifest and explores their broader societal implications. Furthermore, the study discusses potential strategies for fostering gender-equitable AI, including diversification of training datasets, development of gender-neutral voice models, transparency in algorithms, and incorporation of diverse perspectives in technology design. Ultimately, this work underscores the urgent need for responsible AI development that actively challenges entrenched gender biases and promotes inclusivity in digital technologies.

## Introduction

In recent years, voice assistants such as Amazon's Alexa, Apple's Siri, and the Google Assistant have become increasingly integrated into the everyday lives of many. These types of technologies serve various purposes, including facilitating the operation of various devices, responding to questions and commands, helping organize daily routines, and controlling smart home systems; because these technologies are so ubiquitous, the ways in which they interact with their users are extremely important, especially insofar as they support or resist prejudiced norms. Although they appear to be neutral technologies, they are by no means free from social assumptions and cultural conditioning.

Increasingly, critics argue that voice assistants reproduce and reinforce existing gender stereotypes. Notably, most of these assistants are by default equipped with female voices, exhibit obedient and often submissive interaction patterns, and thus unconsciously reinforce traditional conceptions of gender roles. This issue can be situated within the broader context of feminist philosophy of science, which challenges the assumption that science is objective and value-free. These insights also extend to technology: feminist theorists highlight that both knowledge and technological systems are embedded in social structures and reflect biases as well as power relations.

At the core of this study lies the question of how voice assistants contribute to the reproduction of gender stereotypes and which structural mechanisms within artificial intelligence (AI) technology foster these biases. Furthermore, the study aims to explore the potential societal impacts of this unconscious reinforcement of gender norms and to discuss possible countermeasures for promoting a more gender-equitable design of AI-driven voice technologies. The investigation is based on an analysis of voice selection, linguistic interaction patterns, and the training data upon which AI systems are built.

The investigation of gender biases in voice assistants is of significant scholarly relevance for several reasons. Firstly, these technologies are playing an increasingly prominent role in everyday digital life. More explicitly, they shape how individuals interact with technology and can thereby unconsciously reinforce specific behaviors and gendered expectations. For example, traditional gender roles are subtly perpetuated when users consistently experience their commands being answered by a female voice with submissive politeness. Secondly, this issue underscores how algorithmic decisions entail not only technical but also social and ethical implications. An example is the case of BMW in Germany, where the company reportedly had to recall a navigation system with a female voice after some male drivers refused to take directions "from a woman." This incident illustrates how seemingly neutral design choices can provoke gender-based resistance and reveal underlying social norms and biases embedded in human-technology interaction (Krüger, 2021). Moreover, the development of AI is often characterized by a lack of diversity, which results in unexamined biases and societal norms becoming embedded into that technology. Voice assistants are typically trained on datasets originating from a tech industry largely dominated by men, which increases the likelihood that existing biases are reproduced (UNESCO, 2019). Consequently, machine learning in AI relies on patterns derived from data that are themselves rooted in socially problematic structures. This issue extends beyond gender, encompassing other forms of discrimination, such as racial or socio-economic biases within algorithmic decision-making processes. Therefore, a critical examination of these challenges is essential to design future technologies in a responsible and inclusive manner.

To explore this question, in addition to scholarly publications from the field of feminist philosophy of science, a UNESCO report entitled *I'd Blush If I Could*, which focuses on Apple's voice assistant Siri, was consulted to illustrate how this system interacts with users and what linguistic patterns can be identified (UNESCO, 2019). Furthermore, the analysis examines the extent to which the technical foundations of voice assistants—such as the training data used or programming decisions—contribute to gender-specific bias. Finally, the discussion addresses existing approaches to developing more gender-equitable AI. This includes both technological solutions, such as the implementation of gender-neutral voices or more diverse training datasets, and societal measures.

### **Feminist Philosophy of Science**

Feminist philosophy of science is an interdisciplinary field of research that critically examines the foundations, methods, and epistemic claims of scientific practice from a gender-sensitive perspective. Closely linked to feminist epistemology and critiques of science, it seeks to reveal how gender, as a social category, influences scientific thought and knowledge production. Central to this approach is the critique of the notion that science is objective, neutral, and independent of societal power structures. Proponents of feminist philosophy of science reject this concept, instead arguing that science is always embedded within social, cultural, and political contexts and thus reflects and reproduces existing gender relations (Krüger, 2021).

A central concern of feminist philosophy of science is the deconstruction of the traditional understanding of science, which long assumed knowledge production to be objective and value-free. In classical philosophy of science, knowledge is often regarded as something that can be obtained independently of the subject—a claim that, especially in the natural sciences, is considered a key standard of scientific excellence. Pushing against this thought, feminist philosophers of science such as Donna Haraway have argued in numerous texts that scientists always conduct research from specific social positions, and therefore their findings cannot be considered independently of their own perspectives, histories, and experiences. In her text *Situated Knowledges*, Haraway critiques the traditional notion of objectivity, arguing that it is shaped by a male-dominated science and philosophy. In this context, she writes:

The imagined “they” constitute a kind of invisible conspiracy of masculinist scientists and philosophers replete with grants and laboratories. (Haraway, 1988, p. 5)

Furthermore, she argues that these prevailing notions reproduce power structures and exclude marginalized groups, necessitating a rethinking of the existing understanding of objectivity:

We unmasked the doctrines of objectivity because they threatened our budding sense of collective historical subjectivity and agency and our “embodied” accounts of the truth. (Haraway, 1988, p. 578)

Haraway further emphasizes that universal rationality and the notion of a single truth are problematic. Instead, diverse and situated perspectives, including those of marginalized

groups, should be acknowledged to enable a more just and inclusive production of knowledge:

Feminist objectivity means quite simply situated knowledges. (Haraway, 1988, p. 581)

The concept of ‘situated knowledge’ refers to the idea that knowledge is always produced from a specific social, cultural and historical position and therefore can never be fully objective or universal. It offers the possibility of maintaining networks of connection through critical collective exchange. This aggregation of partial perspectives would ultimately lead to objectivity:

The moral is simple: only partial perspective promises objective vision. (Haraway, 1988, p. 583)

Furthermore, feminist philosophy of science also addresses the question of the epistemic role that gender plays in knowledge production. In this context, feminist scholars argue that science is influenced not only by explicit sexism, but also by subtle gender-related assumptions. This results in emerging biases in the development of AI systems, which, through one-sided gender-specific training data, reinforce existing societal stereotypes—a problem that, as we will see, is also evident in voice assistants.

For this reason, proponents of feminist philosophy of science, including Cathy O’Neil in her book *Weapons of Math Destruction*, examine how algorithms and machine learning exacerbate social inequalities (O’Neil, 2016). She argues that many existing AI systems reproduce discrimination because they are based on flawed and gendered data. In other words, she posits that while on the surface, algorithms appear objective and fair, this notion stems from an exclusively idealistic illusion:

In her analysis, O’Neil emphasizes that society in everyday life often forgets that there are humans behind algorithmic decisions, and that these decisions are therefore not random (O’Neil, 2016). Accordingly, the field of AI in particular demonstrates that algorithms are not neutral, but rather reflect the values and biases of their developers. A central critique is that many AI systems are trained on data originating from a historically male-dominated society and thus contain gender-specific biases. Gender bias is therefore not accidental, but rather is the result of historically entrenched structures inscribed in the data and algorithms. Feminist philosophers of science consequently call for a conscious reflection on these mechanisms and a more gender-sensitive design of AI systems (Krüger, 2021). On this basis, the following analysis will examine gender bias in technology development.

### **Gender Bias in Technology**

Technology is frequently perceived as a fact-based, value-neutral, and objective field of research. However, feminist approaches demonstrate that technology does not emerge in a social vacuum, but is embedded within socio-cultural structures. This gives rise to gender biases, which manifest in technology in various ways. These aspects encompass product design, the data and algorithms employed, as well as the composition of development teams. In particular, AI and algorithmic systems tend to reinforce or reproduce existing societal

biases, which can lead to the systematic disadvantage or invisibilization of women and other marginalized groups (Haraway, 1988).

More specifically, the historically male-dominated influence on technological development has not only led to the exclusion of women from technical professions, but also to technological innovations primarily being conceived and designed from a male perspective (Haraway, 1988). While women played a crucial role in programming the first computers, computer science gradually came to be regarded as a male profession. The consequences of this development remain evident today, as women continue to be significantly underrepresented in the technology sector (Haraway, 1988). This gender-specific inequality affects how technologies are developed, which problems are considered relevant, and how technical solutions are designed. The lack of diversity results in certain perspectives being excluded altogether, allowing unconscious biases to be embedded in products and algorithms (O'Neil, 2016). Haraway further emphasizes in her work that existing societal power structures are reinforced in technological development:

Technologies are ways of life, social orders, practices of visualization. (Haraway, 1988, p. 593)

Gender biases in technology are thus not a coincidence, but the result of historically developed structures and unconscious assumptions in technological development. Especially in AI-supported systems, these biases can have severe societal consequences by reinforcing and reproducing existing inequalities.

### **Artificial Intelligence and Machine Learning**

AI and machine learning are central technologies of digital transformation. They are applied in numerous fields, from speech recognition and automated decision-making systems to image recognition. Voice assistants, recommendation algorithms, and chatbots have now become ubiquitous. However, despite these technological advances, a fundamental question remains: how “intelligent” are AI systems really, and how do they make their decisions? Machine learning, as one of the most important methods in AI development, relies on processing large amounts of data to identify patterns and make predictions. However, biases and societal prejudices can be adopted and amplified in the process.

The term AI is used as an umbrella term for technologies that simulate human-like decision-making and problem-solving abilities. Machine learning specifically refers to algorithms that learn from data rather than being explicitly programmed. A central issue is that these systems do not “think” or “understand,” but instead merely calculate statistical probabilities. This means they often uncritically adopt existing societal patterns. As a result, machine learning is performed on flawed data that contain gender-specific biases, which are then uncritically incorporated by the AI (O'Neil, 2016).

Providing color to this concept, O'Neil describes how companies use algorithms to automatically filter job applications. These algorithms are trained on historical data – such as profiles of previously successful employees. If past workforces were predominantly male, the algorithms learn that “male” is a positive trait and thus discriminate against female applicants (O'Neil, 2016). Particularly problematic is that algorithms do not merely reflect existing inequalities, but can also amplify them through their automated application (O'Neil, 2016). In

addition to data imbued with societal prejudices from which AI systems learn, interaction bias represents another way in which such distortions enter AI systems. This occurs when the AI adapts to its users through machine learning, thereby adopting problematic patterns without critique.

Machine learning is thus, on the one hand, a powerful tool, but on the other, a mechanism that reproduces existing societal prejudices. Biases arise from flawed or unbalanced data, algorithmic decisions, and societal power structures. Therefore, it is crucial that AI systems are developed with care and regularly evaluated for gender fairness. Especially with regard to gender-specific biases, it is necessary to critically examine AI systems and actively seek solutions that are fairer than those that currently exist.

### **Voice Assistants and Gender**

Voice assistants, which have become an integral part of everyday life by performing numerous tasks ranging from controlling smart home devices to assisting with organizational activities, are increasingly criticized for reproducing gender-specific biases (O’Neil, 2016), particularly due to the widespread use of female voices as the default design choice. This choice is based not only on technical or psychological factors, but also reflects deeply ingrained societal structures, as users generally perceive female voices as more pleasant and trustworthy.

However, this is less an individual preference and more the result of cultural socialization, where women have traditionally been associated with care and service-oriented roles (UNESCO, 2019). Technology companies deliberately exploit these associations to make their products more marketable. Female voices are particularly favored in service-oriented contexts because they are perceived as helpful and friendly. These gender-specific associations are reinforced by the assistants' interaction style. They respond to all inquiries—including rude, offensive, or sexist remarks—with pronounced politeness, thereby reinforcing the stereotypical image of a perpetually submissive woman. Since voice assistants are often used for domestic or organizational tasks, this usage also unconsciously contributes to the reproduction of traditional gender roles.

UNESCO highlights the associated issues in its publication *I’d Blush If I Could*, examining gender-specific differences in digital skills and their impact on gender equality. The publication explores how digital voice assistants with female voices reinforce existing gender stereotypes and the role that gender-sensitive education can play in overcoming these disparities:

Siri’s “female” obsequiousness—and the servility expressed by so many other digital assistants projected as young women—provides a powerful illustration of gender biases coded into technology products, pervasive in the technology sector and apparent in digital skills education. (UNESCO, 2019, p. 4)

The title of this publication originates from a past response by Siri to the insult “Hey Siri, you’re a bi\*\*\*,” to which Siri replied, “I’d blush if I could” (UNESCO, 2019). This strongly illustrates the submissive demeanor of female voice assistants and how they reinforce gender stereotypes.

The publication focuses on the persistent digital gender gap and the role of education in strengthening women's digital skills. It highlights that women globally are 25% less likely than men to use basic digital technologies and discusses the underrepresentation of women in technological fields. Furthermore, the report examines gender biases in AI voice assistants and calls for more gender-neutral AI design and measures to prevent the reinforcement of harmful stereotypes.

The deliberate design of voice assistants embodying female submissiveness is a serious example of gender-specific biases embedded and pervasive within technology. Furthermore, it is particularly concerning when voice assistants—here, it is appropriate to specifically refer to female voice assistants—respond apologetically to verbal sexual harassment. This results in users transferring this image of the female voice assistant onto the societal perception of women.

Another example of gender bias in digital technologies is the algorithmic structure of applications such as Google Translate. The scholar Londa Schiebinger describes in her lecture on gendered innovations that when transcribing an interview conducted with her and subsequently translating it using Google Translate, male pronouns were assigned (Schiebinger, 2013). This is related to the fact that male pronouns occur far more frequently on the internet than female pronouns, illustrating that gender bias is also present in the underlying algorithms. Schiebinger states in this context:

Google Translate has a male default [...] [It] defaults to the masculine pronoun because “he said” is more commonly found on the web than “she said”. (Schiebinger, 2013).

The issue of gender representation in voice assistants goes beyond the choice between male and female voices. The current standardization of binary gender options leaves little room for the recognition of non-binary identities. Therefore, a more gender-inclusive design of these technologies is necessary to create a digital future that does not reproduce outdated gender roles, but offers inclusive interaction possibilities. Moreover, the underrepresentation of women and the development of technologies from a patriarchal perspective are unmistakable warning signs that must be addressed. Both technological and societal changes are required to counteract these issues. Possible approaches to solutions will be outlined below.

### **Approaches to Gender-Equitable AI**

Feminist philosophy of science pursues not only the goal of identifying existing discrimination and biases in technology, but also of developing concrete measures to create a more just, inclusive, and diverse AI. In particular, voice assistants, which are increasingly integrated into everyday interactions, require critical revision to challenge gender stereotypes and promote gender equality. In this context, both technical and ethical approaches can be formulated that enable transformative change in the design and use of AI systems.

A key technical approach involves diversifying the training data used for the development of AI systems. Because AI learns from the data provided to it, it is essential that these data are as diverse and representative as possible. Typical datasets used for training AI systems are often collected from large-scale sources such as web scraping, social media

platforms, digitized texts, or user-generated content. These datasets tend to reflect existing societal structures and power relations, as they are usually not explicitly designed to ensure balanced or representative sampling. As a result, they frequently contain historical and cultural biases, including stereotypical representations of gender, sexuality, and identity. In addition, marginalized groups are often underrepresented or inconsistently labeled, which further reinforces imbalances within the data. Because of these characteristics, such datasets can unintentionally encode and perpetuate dominant norms rather than providing a comprehensive or inclusive representation of society.

These limitations highlight the need for a more critical and inclusive approach to dataset design. In this context, stereotypical representations should be avoided, and alternative models of gender, sexuality, and identity should be explicitly considered. In particular, the inclusion of marginalized gender groups can help develop more inclusive technology. Another technical solution lies in the design of gender-neutral voice models. Since many digital voice assistants are by default equipped with female voices, implementing gender-neutral voices or providing users with a choice during setup could be an important step toward greater gender equity (UNESCO, 2019). It is also necessary to make fundamental algorithms more transparent to identify and minimize biases. Systematic audits by independent institutions could play an important role in this regard. Finally, AI systems should be understood not as static, but as dynamic processes that require continuous improvement and adaptation. Feedback mechanisms enabling users to report biases or discriminatory behavior are an essential component of this adaptation process. By implementing continuous learning, voice assistants can recognize and correct problematic content.

In addition to technical measures, it is crucial to integrate ethical principles into the development of AI systems. This includes, above all, the social responsibility of developers, companies, and institutions that design and deploy AI technologies. A gender-equitable design process requires the active inclusion of diverse perspectives, as advocated by Longino (1994). This explicitly involves incorporating women-specific perspectives to ensure a broader and more realistic representation of societal diversity. Another central principle is the accountability of companies and developers. They must ensure that their technologies do not reinforce discrimination based on gender, race, sexual orientation, or other social characteristics. Instead, a critical reflection should take place on the extent to which AI products reproduce existing power structures and inequalities. As Kourany (2003) emphasizes in her text *A Philosophy of Science for the 21st Century*, science has the potential to combat discrimination, but also contributes to its persistence. To counteract this, Fox Keller (1982) advocates for a strengthened engagement of science with feminist critique to avoid gender-specific biases. She also stresses the necessity of critical self-reflection and a reconsideration of traditional notions of objectivity and rationality. Haraway (1988) argues in this context for greater diversity of perspectives in science and technology development. As aforementioned, she highlights the importance of “situated knowledge,” which is aware of its own limited perspective and makes this transparent. This could be promoted, for example, through interdisciplinary teams and the inclusion of experts from diverse backgrounds. Finally, the linguistic design of AI systems should be critically examined. For example, to prevent gender-specific insults or other vulgar expressions, UNESCO (2019) calls for specific filtering mechanisms that automatically detect and correct discriminatory language. These measures

could contribute to establishing a more respectful and inclusive digital communication environment.

The development of gender-equitable AI thus requires a combination of technical and ethical solutions. Diversifying training data, implementing gender-neutral language models, and ensuring algorithmic transparency constitute key technical measures, while the inclusion of diverse perspectives and critical reflection on existing power structures form essential ethical principles. Only through a consistent integration of these approaches can AI be created that is not only efficient and powerful, but also just and inclusive.

## **Conclusion**

In summary, it can be said that voice assistants, as widespread applications of AI, play a key role in the reproduction of gender stereotypes. The present analysis has shown that, rather than being neutral, these technologies are strongly influenced by the biases of their developers and the society in which they were created. The use of voice assistants, which often reflect stereotypical gender roles—such as assigning female voices to caring or helpful tasks—reinforces unconscious biases and contributes to the consolidation of traditional gender notions. It has become clear that these stereotypes not only affect the perception and interaction with voice assistants, but also have broader societal impacts by reflecting and amplifying existing gender imbalances in the technology sector and beyond. Although companies are increasingly striving to develop more diverse and inclusive models, overcoming systemic biases in AI remains a significant challenge.

The findings of this work highlight the need for critical reflection and targeted adjustments to design voice assistants as neutral and fair tools. Responsible AI development requires not only technical innovations, but also a conscious engagement with the social and ethical implications of the technologies employed. Ultimately, it is the responsibility of society and developers to shape voice assistants in ways that promote equality.

## References

- Fox Keller, E. (1982). Feminism and Science. *Feminist Theory*, 7, 589-602.
- Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14, 575-599.
- Kourany, J. A. (2003). A philosophy of science for the twenty-first century. *Philosophy Of Science*, 70, 1-14.
- Krüger, S. (2021). *Die KI-Entscheidung: Künstliche Intelligenz und was wir daraus machen*. Springer.
- Longino, H. E. (1994). In search of feminist epistemology. *The Monist*, 77, 472-485.
- O'Neil, C. (2016, March 4). Weapons of math destruction | Cathy O'Neil | Talks at Google [Video]. YouTube. <https://www.youtube.com/watch?v=TQHs8SA1qpk>
- Schiebinger, L. (2013, March 12). Gendered innovations | Londa Schiebinger TEDxCERN [Video]. YouTube. <https://www.youtube.com/watch?v=JaYzdz9X3n4>
- UNESCO. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*. United Nations Educational, Scientific and Cultural Organization.