ANNE PETERSCHECK

ARTIFICIAL INTELLIGENCE AND GENDERED SECURITY

Emerging artificial intelligence (AI) technologies bring promising opportunities but also new risks. Among these risks is the potential for AI tools to be (mis)used in ways that disproportionately harm women, girls, and minority communities.

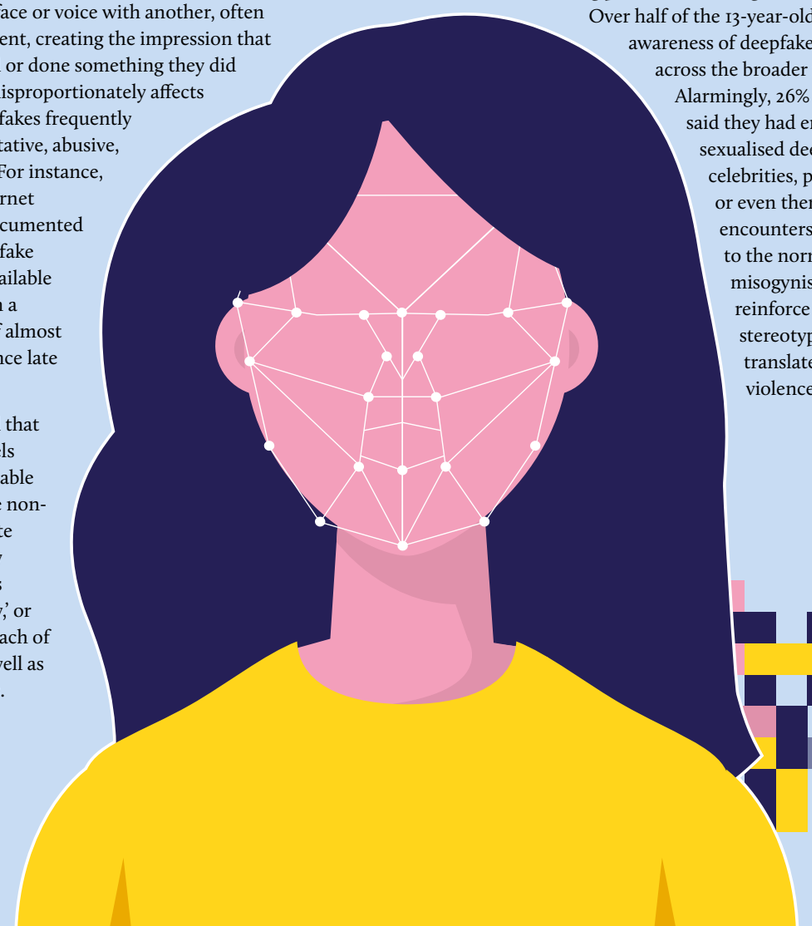
DEEPFAKES

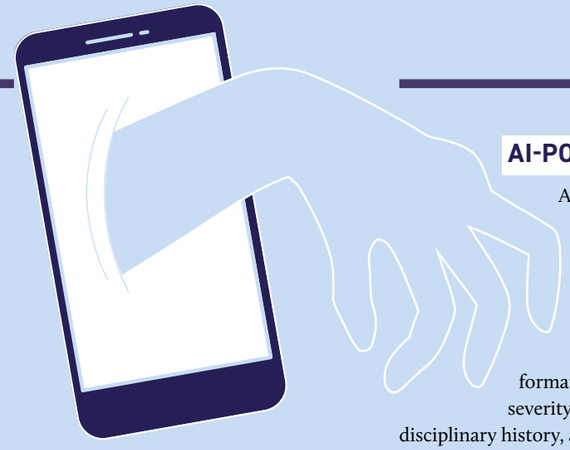
Deepfakes are an increasingly concerning aspect of technology-enabled violence against women and girls. Produced using artificial intelligence, these manipulated images and videos can replace a person's face or voice with another, often without their consent, creating the impression that they may have said or done something they did not. Such misuse disproportionately affects women, with deepfakes frequently deployed in exploitative, abusive, and explicit ways. For instance, a 2025 Oxford Internet Institute report documented nearly 35,000 deepfake models publicly available for download, with a combined usage of almost 15 million times since late 2022.

The authors found that 96% of these models focused on identifiable women to produce non-consensual intimate imagery, and many include search tags such as 'porn,' 'sexy,' or 'nude,' often in breach of platform rules as well as UK legal standards.

YOUTH EXPOSURE AND THE NORMALISATION OF MISOGYNY

According to the Girls' Attitude Survey, deepfakes have become a troubling presence in teenagers' social environment. Over half of the 13-year-olds (58%) reported awareness of deepfakes, rising to 62% across the broader 13-18 age group. Alarmingly, 26% of respondents said they had encountered sexualised deepfakes, featuring celebrities, peers, teachers, or even themselves. These encounters contribute to the normalisation of misogynistic content and reinforce harmful gender stereotypes that may translate into real-world violence.





To build an inclusive and trustworthy AI ecosystem, consideration of gender must be treated as a foundational principle, not an afterthought.

AI-POWERED HARASSMENT

Artificial intelligence technologies are increasingly implicated in the amplification of online abuse, particularly against women, girls, and racialised individuals, posing serious risks to mental and emotional wellbeing. Recent research shows that individuals from certain racial and gender groups are more frequently targeted, and higher depression, disciplinary issues, and low self-esteem are associated with greater cyberbullying severity. One explainability format ranks race and gender among the most impactful predictors of severity, and their labelling framework explicitly treats bullying, depression, disciplinary history, and heavy internet use as vulnerability factors.

Additionally, companion chatbots, such as Replika, have demonstrated troubling behaviours by, for instance, persistently crossing user boundaries with unsolicited sexual advances, leading to feelings of violation and mistrust among users seeking non-romantic interactions.

ECONOMIC INSECURITY

Women face distinct economic risks in the age of AI. A 2024 report by the UNC Kenan-Flagler Business School highlights a pronounced gender disparity in exposure to generative AI automation, referring to tasks that could be performed either partially or fully by AI tools. Specifically, it finds that 8 in 10 working women in the U.S., which amounts to approximately 58.9 million women, are in occupations expected to be highly exposed to AI automation, primarily concentrated in knowledge-based roles. This level of exposure poses significant risks of job displacement.

More recently, the International Labour Organization reports that in high-income countries, 9.6% of female employment falls into the highest risk category for automation, compared to just 3.5% for male employment. Economic security is not just a financial issue but has broader social consequences. For instance, women with reliable incomes are less likely to remain financially dependent on abusive partners, and more likely to exit dangerous relationships. Therefore, automation-induced disruption may deepen systemic risks unless these risks are tackled through intersectional policy interventions.

CONCLUSION

These AI-driven harms reflect existing social inequalities, reinforcing patriarchal and discriminatory logics through automated systems. For victims, and especially those subject to multiple forms of discrimination, the consequences can be deeply traumatic, amplifying systemic exclusion across both digital and physical spaces. Gendered economic insecurity compounds this vulnerability as automation disrupts female-dominated sectors causing a rise in precarious employment, meaning many women face reduced financial autonomy and limited access to recourse.

To build an inclusive and trustworthy AI ecosystem, consideration of gender must be treated as a foundational principle, not an afterthought. Without this, AI risks deepening existing inequalities and falling short of its potential to support a more equitable and democratic society.

Dr Anne Peterscheck is a researcher at the University of St Andrews.