PUBLIC SPHARE:
Jurnal Sosial, Politik, Pemerintahan dan Hukum
ISSN (Online): 2807-3878
DOI: https://doi.org/10.59818/jps.v5i3.3293
Vol. 5, No. 3, September 2026

IMPLICATIONS OF AI GROK-BASED NCII PRODUCTION ON FEMALE BRAWIJAYA UNIVERSITY STUDENTS' PERSONAL SECURITY ON SOCIAL MEDIA

Azahra Aulia Ramadani[1], Asih Purwanti[2]
Brawijaya University, Malang, Indonesia
email: aazahraauliar@gmail.com[1], a.purwanti@ub.ac.id[2]

RIWAYAT ARTIKEL
Received : 2026-09-14
Revised : 2026-10-23
Accepted : 2026-11-02

KEYWORDS
artificial intelligence; human security; personal security; image-based sexual abuse; non-consensual intimate images

KATA KUNCI
kecerdasan buatan; keamanan manusia; keamanan personal; kekerasan seksual berbasis gambar; NCII

ABSTRACT
Generative AI such as Grok has accelerated the production of Non-Consensual Intimate Images (NCII) on platform X, a form of gender-based sexual violence threatening women's personal security, yet its effects on victims' subjective sense of safety remain underexamined. This study examines how AI Grok-based NCII production affects the personal security of female Brawijaya University students across five human-security dimensions: psychological security, freedom from fear, dignity, social participation, and protection-empowerment capacity. Using a qualitative descriptive design, data were gathered from 23 purposively sampled respondents via an open-ended online survey and literature review, analyzed thematically. Results show NCII is perceived as a threat extending beyond privacy, generating anxiety and anticipated insecurity even without direct victimization, reputational and dignity harm tied to gendered stigma, self-censorship restricting digital participation, and persistent distrust toward institutional protection despite some confidence in AI-based detection tools. It is concluded that AI-based NCII constitutes a multidimensional personal-security threat requiring accessible reporting, digital safety education, responsive moderation, and institutional support.

ABSTRAK
AI generatif seperti Grok telah mengakselerasi produksi Non-Consensual Intimate Images (NCII) di platform X, sebuah bentuk kekerasan seksual berbasis gender yang mengancam keamanan pribadi perempuan, namun dampaknya terhadap rasa aman subjektif korban masih belum banyak diteliti. Penelitian ini mengkaji bagaimana produksi NCII berbasis AI Grok memengaruhi keamanan pribadi mahasiswi Universitas Brawijaya melalui lima dimensi keamanan manusia: keamanan psikologis, kebebasan dari rasa takut, martabat, partisipasi sosial, serta kapasitas perlindungan dan pemberdayaan. Menggunakan desain deskriptif kualitatif, data dikumpulkan dari 23 responden yang dipilih secara purposive melalui survei daring terbuka dan kajian literatur, lalu dianalisis secara tematik. Hasil menunjukkan NCII dipersepsikan sebagai ancaman yang melampaui pelanggaran privasi, menimbulkan kecemasan dan ketidakamanan antisipatif meski tanpa viktimisasi langsung, kerugian reputasi dan martabat yang terkait stigma gender, swasensor yang membatasi partisipasi digital, serta ketidakpercayaan yang menetap terhadap perlindungan institusional meski ada sedikit keyakinan pada alat deteksi berbasis AI. Disimpulkan bahwa NCII berbasis AI merupakan ancaman multidimensional terhadap keamanan pribadi yang memerlukan mekanisme pelaporan yang mudah diakses, edukasi keamanan digital, moderasi yang responsif, dan dukungan institusional.

1. Introduction

The digital realm has changed how people interact globally, with the ease of real-time communication with a wider audience affording greater freedom to express themselves, particularly through social media. Although social media is often used as a medium for women to express themselves freely, the content is often duplicated and misused without consent for various crimes, such as *online stalking*, misuse of information for blackmail, intimidation, and reputational damage. This can be seen as ironic, as media that should be a source of freedom for women can instead be used as a mechanism that facilitates *online sexual violence* (Lakshané, 2024).

The development of digital technology has shaped a new global public space by expanding global social interactions while giving rise to new forms of gender-based sexual violence, including *Non-Consensual Intimate Images* (NCII). This article reflects how the landscape of sexual violence has shifted from physical forms to the ecosystem of *Non-Consensual Intimate Images* (NCII) in the digital world. This paper examines how the integration of AI (particularly GROK) facilitates the production of NCII, which threatens women's personal security worldwide.

Non-Consensual Intimate Images (NCII), which refers to the distribution of intimate digital material without the consent of the individuals depicted, are one example of how AI can produce and distribute content in a more realistic, faster, and more easily replicable manner. This has been demonstrated in recent years by the integration of generative AI into numerous social media platforms. The definition of NCII varies widely and is largely influenced by the social and cultural norms and settings of many countries worldwide. NCII adversely affects women's dignity and restricts their freedom by violating their right to privacy, sexual consent, and freedom from abuse and fear. (Lakshané, 2024).

The ability of generative AI to perform *deep learning* enables the production of highly realistic and detailed synthetic intimate content without directly involving individuals (commonly known as *Deepfakes*). However, it still causes significant reputational and psychological damage. The production, customization, and distribution of visual content are accelerated by using GROK generative AI on the digital platform "X" (formerly Twitter) and other generative AI media. This complicates the distinction between a digitally manipulated product and reality (Chesney & Citron, 2018). This phenomenon cannot be separated from changes in the structure of public space. As the internet has become part of human social life, the distinction between the *online* and *offline* worlds has become increasingly blurred. However, digital space continues to facilitate the *digital divide* that makes women more vulnerable to gender-based violence, especially given that the anonymity and deindividualization often offered in social media make *online* social spaces a strategic hunting ground for *online predators* (Organization of United States, 2017)

Recent developments show that the integration of artificial intelligence into social media platforms is bringing about major changes in the production and distribution of digital content. AI systems can function as moderators and recommenders, as well as content generators, accelerating the spread of visual representations and digital narratives worldwide. In this situation, NCII is no longer merely an individual act; it now operates within a complex digital ecosystem comprising global platforms, algorithms, anonymity, and cross-border distribution networks. In many cases, these are not only symbolic but also affect freedom, social mobility, and even women's involvement in the public sphere (Salerno-Ferraro *et al.*, 2021). This condition makes women more vulnerable to personal security threats in the digital world.

In this context, this paper examines how integrating GROK, a generative AI technology, can facilitate the production of NCII and its impact on female Brawijaya University students.

2. Literature Review

The concept of personal security is one dimension of human security first introduced by the UNDP in the 1994 *Human Development Report*. According to the UNDP report (1994), security studies have traditionally focused on protecting states or regions from external threats. Human security itself is based on two aspects. First, security against life-threatening threats, such as hunger, disease, and oppression. Second, security from incidental threats to freedoms that disrupt daily life, whether individually, in groups, or in the workplace. Thus, it can be said that human security is a condition in which society or an individual is free from fear and deprivation.

In the global context, security often refers to maintaining the stability of the system of states. In security studies, the state is a vital actor that must be protected from external threats. However, it is not uncommon for the public to fall victim to external threats or even to be threatened by their own

government's policies. The concept of personal security is grounded in an approach focused on protecting human life (Andersen-Rodgers & Crawford).

In this context, a state of safety essentially occurs when a person no longer feels pressured by the potential for loss. Once they become aware that something might slip from their grasp, humans subconsciously shift their priorities to address it. The problem is that attention to this risk often becomes overwhelmingly dominant. People typically prioritize risk management when given the option between resolving that concern and maintaining attention on other objectives. People may give up many other things to lessen this fear if it persists. Long-term cognitive decline and poor decision-making may result from this, possibly causing psychological anguish (Arkin *et al.*, 2015).

The phenomenon of *Non-Consensual Intimate Images* (NCII) shows how personal security can collapse severely if it is regarded as a state of being fearless. NCII poses a persistent and tangible threat to victims' mental health, reputation, and social relationships due to the potential for intimate content to be disseminated without consent. Additionally, NCII is frequently employed as a weapon for hegemonic masculinity construction and symbolic dominance, with offenders exploiting such content to manipulate and degrade female victims. (Im *et al.*, 2022; Viola & Voto, 2023; Almeida *et al.*, 2026). This issue is not limited to a small scale but involves a significant number of cases.

The generation of NCII has greatly grown as generative artificial intelligence (AI) technology advances. According to research by Hawkins *et al.* (2025), there has been a notable annual growth in the quantity of *deepfake* photos. Such crimes frequently put women in a precarious situation. Technology-enabled sexual violence disproportionately targets women in various forms, such as *cyber harassment*, image-based sexual abuse, and NCII (Maras & Logie, 2024). This impacts not only adults but also adolescents and even children. In the same study, 15% of high school students in the United States had at least heard of cases involving sexually explicit synthetic visual content that included other students at their school. Nevertheless, women remain the largest target group, with a rate reaching 99% (Hawkins *et al.*, 2025).

Women are the group that is most vulnerable to gender based violence in the digital sphere. A study by Gupta and Gogoi (2025) shows that women are more frequently targeted and experience more serious consequences, such as psychological distress, physical threats, and restrictions on their participation in the digital world. Additionally, research conducted by IM *et al.* (2023) indicates that women are more fearful of digital threats, particularly the dissemination of NCII, which directly impacts their safety and social reputation.

To date, research on personal safety, gender-based violence, and NCII has advanced significantly. However, most research still focuses on conventional NCII or user-generated content, while research on NCII generated by generative AI remains very limited. In fact, *text-to-image* and *deepfake* technologies have advanced rapidly and significantly reduced technical barriers, enabling the production of NCII with fewer resources and in a shorter time (Hawkin *et al.*, 2025). Furthermore, very few studies specifically incorporate personal security aspects into their approaches to examining this phenomenon. Most research emphasizes legal or technological aspects (Furizal *et al.*, 2025). It does not link them to the victims' subjective experiences, such as their sense of safety, self-confidence, and their level of engagement in the digital world.

3. Method

This study employs a qualitative, descriptive research design to explore individuals' perceptions and experiences of *Non-consensual Intimate Image* (NCII). This method was chosen to investigate how people form social realities and attach subjective meanings to their experiences, particularly regarding concerns about personal safety in digital spaces. (Lamont, 2015). Additionally, this approach is frequently used in social studies to examine problems that are based on experience, identity, and perception and cannot be quantified (Creswell, 2014; Hannes *et al.*, 2022).

This study explores the perceptions of female students at Brawijaya University regarding *Non-Consensual Intimate Images* (NCII) generated by artificial intelligence on the social media platform X. It focuses on individual experiences, risk assessments, and the strategies these students use to mitigate the associated threats, specifically targeting those who frequently use X, previously known as Twitter.

Primary data will be collected through an *online*, open-ended survey using Google Forms, enabling the researcher to obtain detailed contextual information about informants' experiences and understanding of the phenomenon under study (Kvale & Brinkmann, 2009). Secondary data for the study were collected via a literature review consisting of research papers and academic articles

on NCII, AI, and digital security. Integration of primary and secondary data through triangulation enhances the validity of conclusions and enriches the research (Patton, 2002).

This study utilized both primary data from interviews and secondary data from academic literature. A case study approach employing thematic analysis was then applied to analyze the data, enabling the identification and interpretation of patterns within the qualitative data (Heriyanto, 2018). This method is frequently employed in social science research and is considered helpful for understanding individual experiences and perceptions within complex social contexts (Nowell *et al.*, 2017). Several steps are involved in the analysis process, including becoming familiar with the data, coding, organizing codes, and creating themes relevant to the research question.

Purposive sampling will be utilized in this study to select informants who align with the research goals. This method emphasizes gaining a deep understanding of the issue rather than seeking statistical generalizations. Female students at Brawijaya University who are active on social media and have knowledge of or opinions on digital security and the misuse of AI-generated content will be selected as informants. Initial informants may also refer additional relevant participants, considering the sensitivity of the topic.

This research utilizes data saturation, determining that data collection ceases when no new significant information emerges. The number of informants ranges from 5 to 20, depending on the depth of the data. A homogeneous sampling strategy is employed, focusing on students while accounting for variations in experience and digital literacy. This approach aims for rich, credible, and sharp analytical data, showcasing the complexity of social experiences without compromising interpretive validity (Miles *et al.*, 2014).

By emphasizing personal security, the study applies the concept of human security as described in the 1994 United Nations Development Program *Human Development Report*. This perspective expands the notion of security beyond state defense against external threats to include the protection of individuals from a range of dangers that threaten their safety, dignity, and freedom (United Nations, 1994). In this context, personal security refers to the protection of individuals from physical, psychological, social, and structural violence perpetrated by the state, social groups, or other individuals, including threats that specifically target women.

Based on this framework, the implications of the AI-based NCII Grok for the personal security of female students at Brawijaya University will be examined through five dimensions derived from the concept of personal security within the human security framework. The first indicator is psychological security, which refers to a psychological state free from anxiety, fear, and emotional stress. This indicator is based on the concept of personal security developed by Arkin *et al.* (2015) and on the findings of Almeida *et al.* (2026) regarding the psychological effects experienced by victims of NCII, which are used to reinforce it.

The second indicator encompasses freedom from fear, which is a person's freedom from the fear of harm that could threaten their safety and well-being. This concept of human security was proposed by the UNDP (1994) and elaborated by Andersen-Rodgers and Crawford (2023), who identify freedom from fear as a key component of personal security.

Next is the protection of human dignity, which means protecting a person's dignity, privacy, and integrity from actions that degrade or exploit them. The human security approach developed by the UNDP (1994) forms the basis for this indicator. Meanwhile, the characteristics of NCII are associated with violations of human dignity and bodily autonomy, as explained by Viola and Voto (2023) in their research.

The fourth indicator is social participation and autonomy, which refers to a person's ability to interact and express themselves freely in social settings without being threatened or restricted. This indicator aligns with the personal security dimension developed by Arkin *et al.* (2015). It is supported by findings from Im *et al.* (2022) on women's tendency to self-censor and restrict their *online* activities after experiencing digital harassment.

The final indicator encompasses protection and empowerment capacity, which refers to the ability of individuals and institutions to provide protection and strengthen capabilities to address threats. This indicator is derived from the protection and empowerment components of UNDP's (1994) human security framework. It is reinforced by the findings of Maras and Logie (2024) on the importance of *multi-stakeholder* protection mechanisms in addressing the threat of sexual *deepfakes*.

4. Results and Discussions

According to a thematic analysis conducted on 23 respondents, a common pattern emerged: non-

consensual intimate images (NCII) generated by AI are viewed as a threat that not only involves a violation of digital privacy but also extends to a threat to individuals' psychological well-being, dignity, social participation, and their ability to protect themselves in the digital world. Almost all of the participants were aware that AI technology could be abused to alter someone's identity or photographs. Respondents acknowledged that generative AI technology makes it feasible to create fake content that appears intimate, and viewed NCII as the sharing of private pictures without consent. According to this perspective, the characteristics of digital threats have changed. People do not need to produce or store intimate photographs to become targets because visual material that was initially non-intimate can be converted into sexual content without the subject's consent.

These findings align with the notion of image-based sexual assault provided by McGlynn and Rackley (2017). According to this theory, the creation and distribution of explicit photos without permission is technology-based sexual violence. They emphasize that in addition to losing their privacy, dignity, and sexual autonomy, victims of such treatments experience psychological and social repercussions. As a result, respondents' perception that AI-generated NCII constitutes a serious violation of privacy and personal integrity might be seen as an admission that personal security is at risk from digital threats. This aligns with the UNDP's human security approach, which prioritizes the protection of individuals, their lives, and their dignity over states and territories (United States, 1994).

Psychological security is the first and most obvious characteristic to appear in the data. Nearly all respondents used terms like "fear," "anxiety," "trauma," "shame," "discomfort," "stress," and "insecurity" to describe AI-generated NCII. Even individuals who had never been direct victims experienced concern when they realized that pictures exchanged for non-sexual reasons could be transformed into intimate content. One respondent said they would be "anxious, afraid, and certainly traumatized" if their digital identity or photos were changed. Another respondent claims that using social media is not entirely safe due to the possibility of photos or personal identities being misused. Respondents believe that Grok's presence creates psychological discomfort even before actual abuse. This is significant because, from the standpoint of personal security, people lose their sense of security when there has been violence as well as when they are constantly in danger.

These findings indicate that the impact of NCII on personal security can emerge even before actual victimization occurs. Put another way, the dread of becoming a victim has created psychological suffering as a result of the perceived threat. From the standpoint of human security, this is important since security depends on both the lack of violence and the circumstances in which people live in constant fear of danger. The UNDP relates human security to vulnerability and psychological insecurity, and identifies freedom from fear as one of its pillars (United Nations, 1994).

Research on technology-facilitated sexual violence against women supports these empirical findings even more. Champion *et al.* (2022) show that technology-facilitated sexual violence can have an impact on victimization associated with experiences of gender-based violence in digital settings. Additionally, a recent systematic analysis by Benítez-Hidalgo *et al.* (2026), which integrated 21 studies from eleven countries, discovered a connection between depression, anxiety, PTSD, stress, and negative body image and technology-facilitated sexual violence. As a result, the respondents' concern is more than just a personal emotional response; it is proof that the danger of NCII can disrupt the psychological components of women's personal security.

Freedom from fear is the second dimension. According to the findings, respondents frequently associate the use of AI with a higher risk of privacy violations. The loss of control over who can access or use one's digital content is considered the most perilous scenario, and many individuals become aware of their digital identity when personal images are published or shared. Although some claim that private accounts provide greater security, they nevertheless acknowledge that this does not completely remove the possibility of account abuse.

These findings draw attention to a paradox in digital security. The respondents think social media usage is still reasonably secure. However, this security is just temporary and relies on your alertness. Therefore, security is viewed as a condition that must be continuously maintained through access restrictions, content selection, and digital identity management. This situation indicates that NCII generated by artificial intelligence causes a sense of anticipated insecurity, that is, a sense of insecurity stemming from the reality that a person can lose control over their self-representation at any time.

According to UNDP, freedom from fear means a state in which people can live without threats to their safety or freedom (UNDP, 1994). Because female users must constantly consider whether the photos they upload might contain sexual content, these threats have disrupted their ability to use digital spaces. Therefore, NCII is not merely a threat at the level of "harmful content," but has become part of social media users' everyday experience of security.

Furthermore, the context of Grok is important because this threat is not entirely hypothetical. In 2026, xAI itself explicitly recognized the category of non-consensual intimate content, which includes synthetic content created or modified using AI, and provided a notice-and-removal mechanism for reporters (SpaceXAI Creators of Grok, the AI Chatbot, n.d.). This reinforces the relevance of respondents' perceptions that generative AI can indeed serve as a medium for producing NCII. However, respondents' perceptions of Grok's specific capabilities and risks should be distinguished from the technical facts about its features.

The third dimension is the protection of human dignity. Essentially, all respondents viewed the creation of fake content without consent as a serious act because it can damage a person's privacy, reputation, dignity, and integrity. Respondents repeatedly stated that the social impact of AI-generated content remains real even if the content is fake. Some respondents cited the potential for stigma, gossip, humiliation, judgment, and damage to a victim's reputation. This perception suggests that the harm caused by NCII does not depend on the material's quality. The main issues are the absence of personal control over one's own sexual portrayal and the social fallout from its spread.

For Brawijaya University students, this threat to dignity takes on an additional social dimension due to the close-knit, repetitive campus setting. Respondents stated that victims may become the subject of gossip, experience verbal abuse, intimidation, rejection, or even choose to withdraw from college life. As a result, NCII may cause reputational damage, leading to societal suffering. This result is in line with a study conducted in Indonesia on *deepfake pornography*, which revealed that female victims may suffer from psychological trauma, a loss of confidence, social exclusion, and stigma because of a damaged reputation. (Kasita, 2022).

These findings are also highly relevant to the concept of *image-based sexual abuse*. According to McGlynn and Rackley (2017), the consequences of this type of violence include the loss of sexual autonomy, privacy, and dignity. Additionally, McGlynn and Rackley (2017) situate this practice on a continuum of sexual violence, indicating that the problem should be viewed as more than just a transgression of *internet ethics* or a technological abuse. In the context of this study, AI manipulation increases the spectrum of sexual assault; therefore, the problem should be viewed as more than just a transgression of *online ethics* or technological abuse. Because victims do not even need to process or have ever created authentic intimate content, AI manipulation in the context of this study expands this continuum. The technology allows a person's visual identity to be sexually reconstructed without consent.

AI-based NCII "does not affect a person's dignity or reputation, because AI-generated content is merely manipulation," according to an intriguing response. This minority opinion demonstrates that views on NCII are not entirely uniform, which makes it methodologically relevant. This viewpoint, however, is at odds with mainstream opinion, which holds that inaccurate information can still have significant social repercussions.

Additionally, a gender dimension emerged from the respondents' responses. They believed that women were particularly vulnerable because bogus sexual content can result in stigma and more unfavorable societal judgment. These results are consistent with technology-facilitated sexual violence, which shows that digital violence has a significant gender component and disproportionately affects women. One form of technology-facilitated sexual violence against women is *image-based sexual violence* (Henry *et al.*, 2020). Because it simultaneously exacerbates gender inequality in the digital environment, NCII produced by artificial intelligence might be regarded as a threat to individual safety.

The fourth dimension is social participation and autonomy. In this dimension, the impact on respondents' digital behavior was very evident. Every participant stated that AI-generated NCII can, at the very least, affect their freedom to use social media. Most respondents stated that they are afraid to upload photos because they worry that AI or other parties may use the photos. Other changes include being more cautious when uploading photos, reducing the sharing of personal information, making accounts more private, and limiting interactions with strangers.

One could consider this trend a form of *self-censorship*. Although people can still access digital areas, their true freedom is restricted by the threat. Responses about women's self-confidence make this even more evident. Some respondents also connected these dangers to obstacles that prevent them from sharing images, attire, actions, or facets of their personalities on social media.

A recent study on non-consensual *deepfakes* supports these results. According to a 2026 study of 423 female South Korean users, awareness of *deepfake* dangers was positively associated with *online self-censorship*, as evidenced by increased passive social media use and decreased perceived control (Lee, 2026). According to the study, women's engagement and expression in digital spaces may decline due to the threat of *deepfakes*. Therefore, rather than merely a propensity to be "more cautious," the patterns identified in data from Brawijaya University students might be understood as a limitation of digital sovereignty stemming from AI-based sexual risks.

It is interesting to note that the level of limitation varies. Not every respondent thought that their digital identities needed to be entirely altered. Some claimed they could manage risk through privacy settings and content selection without changing who they were. Others stressed the necessity of using pseudonyms or anonymous identities, or of limiting the sharing of personal data. According to the "homogeneous sampling with internal variation" approach (Patton, 2014; Creswell, 2014), this variation indicates internal variation in respondents' experiences. Respondents nevertheless exhibit varying degrees of digital literacy, risk perception, and adaptability techniques, even though they all share the trait of being students.

The fifth dimension is the capacity for protection and empowerment. The figures show a strong conflict between high risk awareness and low confidence in defensive abilities. The majority of respondents stated that there are still insufficient precautions in place by the government, on campus, and through individual actions. Responses such as "not yet," "not enough," "insufficient," and "need further improvement" were commonly employed. Some respondents even admitted that they did not know what to do if they were victims.

Questions about reporting practices brought this disparity to light even further. Out of the 23 respondents, very few had a thorough understanding of the reporting procedures. Most others only knew that victims may report to the platform, the institution, or relevant agencies; they had no idea about the exact procedures. Therefore, user trust, comprehension, and accessibility are essential factors influencing safety, in addition to the existence of formal systems.

These findings are in line with research by Huber & Ward (2025), which shows that the effectiveness of NCII removal depends on victims' familiarity with a range of platforms and reporting processes. This study emphasizes the importance of making the reporting process more accessible because victims' knowledge of the various mechanisms is critical to the success of NCII.

The findings show that respondents are not entirely passive when presented with risks, even though they believe institutional safeguards remain inadequate. Making accounts private, limiting photo uploads, using pseudonyms or anonymous identities, and exercising greater caution before posting anything are a few of the new tactics. Uncertainty about platforms' and organizations' ability to prevent abuse is the driving force behind the inclusion of individual risk management in this strategy.

However, this tactic yields contradictory results. Enhancing personal security can reduce risk exposure, but it may also limit one's ability to interact with the digital world. The strong link between responses to self-defense strategies and responses to freedom-of-expression issues demonstrates this. Respondents begin considering whether to post information, upload photos, or appear in public on social media, as well as performing technical activities such as changing privacy settings. Therefore, protective measures that place the entire cost on individuals can shift the burden of prevention from offenders and technology providers to potential victims.

This viewpoint is crucial because individual vigilance alone is insufficient to ensure personal security. Effective security policies must involve a broader range of actors, including governments, educational institutions, artificial intelligence companies, social media platforms, and the general public, as the UNDP places a strong emphasis on protecting human life, dignity, and freedom.

It is interesting to note that the research data shows more than just unfavorable opinions on AI. Artificial intelligence has the potential to identify, stop, and eliminate NCII, according to most respondents. They recommend employing digital hashing, content moderation, automatic alerts, *deepfake* detection, image-manipulation detection systems, blocking of dangerous content, and faster content removal techniques. The findings show that respondents do not consider technology to be a threat

in and of itself. Instead, they view technology as a tool with numerous advantages that rely on how it is created, utilized, and managed.

This viewpoint is crucial because AI can detect manipulation and enhance moderation by creating synthetic content. However, victim-centered usage of the technology is required. Huber and Ward (2025) show that the platform's technical capabilities and victims' ease of using reporting methods affect the efficacy of NCII removal. Thus, the development of reporting systems that are transparent, *user-friendly*, and sensitive to victims' needs must be integrated with AI-driven protection.

The fact that xAI has a formal reporting system for synthetic NCII in the Grok setting demonstrates that institutional and technical protections can be implemented. Synthetic NCII content is illegal, according to xAI, which also offers reporting and removal processes (SpaceXAI Creators of Grok, the AI Chatbot, n.d.). The study's field results, however, show that formal methods do not always translate into a sense of security among student users. This discrepancy between accessibility and availability seriously threatens personal security.

5. Conclusion

This study examines how 23 female students at Brawijaya University who actively use X perceive the implications of AI-based non-consensual intimate images (NCII) for their personal security. The findings, analyzed through five dimensions of personal security within the human security framework, indicate that the perceived threat extends beyond privacy concerns to include psychological security, freedom from fear, human dignity, social participation and autonomy, and the perceived capacity to obtain protection and support. Respondents expressed concerns about the potential misuse of their images, reputational harm, and the possibility that such risks could restrict their self-expression and participation in digital spaces. They also perceived existing protection and reporting mechanisms as insufficiently accessible or effective.

The study's main contribution is to demonstrate how perceived risks associated with AI-based NCII can be examined as a personal security concern through users' subjective experiences. The findings suggest that the anticipation of image misuse may influence women's digital behavior even without reported direct victimization. This highlights the importance of considering not only the occurrence of harmful content but also the ways in which perceived threats may affect autonomy and participation in *online* spaces. In practical terms, the findings point to the need for further attention to accessible reporting mechanisms, digital safety education, responsive platform moderation, and institutional support. These implications should be understood as areas for policy and institutional consideration rather than as interventions whose effectiveness has been established by this study.

Several limitations should be acknowledged. The study involves a relatively small, context-specific sample of 23 female students from one university, and its findings reflect respondents' perceptions rather than the prevalence of NCII or the experiences of all potential victims. Furthermore, the study does not independently test Grok's technical capabilities to generate NCII. Future research could compare perceptions across universities and social contexts, examine the experiences of individuals who have experienced *image-based sexual abuse* while ensuring appropriate ethical safeguards, and evaluate the accessibility and effectiveness of platform reporting and content-removal mechanisms. Such research would help clarify how perceived and experienced digital threats affect women's personal security and what forms of institutional protection are most effective.

6. References

Almeida, T. C., Cardoso, J., Caridade, S., Pires, A. R., Matos, A. F., Costa, M., & Cunha, O. (2026). Mental health on victims of non-consensual intimate image-sharing: A scoping review. *Sexuality Research and Social Policy*. https://doi.org/10.1007/s13178-026-01312-x

Andersen-Rodgers, D., & Crawford, K. F. (2023). *Human security: Theory and action* (2nd ed.). Rowman & Littlefield.

Arkin, R. M., Wichman, A. L., & Carroll, P. J. (2015). Introducing handbook of personal security. In *Handbook of personal security* (pp. 1–17). Psychology Press. https://doi.org/10.4324/9781315713595

Benítez-Hidalgo, V., Pastor-Moreno, G., Ruiz-Pérez, I., & Amaya-Santos, S. (2026). Technology-facilitated sexual violence against women and its health impact: A mixed-methods systematic review. *BMC Public Health, 26*, 543. https://doi.org/10.1186/s12889-026-26260-4

Champion, A. R., Oswald, F., Khera, D., & Pedersen, C. L. (2022). Examining the gendered impacts of technology-facilitated sexual violence: A mixed methods approach. *Archives of Sexual Behavior, 51*(3), 1607–

1624. https://doi.org/10.1007/s10508-021-02226-y

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE.

Gupta, D. N., & Gogoi, S. (2025). Cyberbullying and gender: An analysis of legal protections for women on social media. *International Journal of Environmental Sciences, 11*(19).

Hannes, K., Hendricks, L., Brgles, M., & Dierckx, C. (n.d.). *The SAGE handbook of qualitative research design: Being creative with resources in qualitative research*.

Hawkins, W., Russell, C., & Mittelstadt, B. (2025). *Deepfakes on demand: The rise of accessible non-consensual deepfake image generators* (arXiv:2505.03859). arXiv. https://doi.org/10.48550/arXiv.2505.03859

Henry, N., McGlynn, C., Flynn, A., Johnson, K., Powell, A., & Scott, A. J. (2020). *Image-based sexual abuse: A study on the causes and consequences of non-consensual nude or sexual imagery*. Routledge. https://doi.org/10.4324/9781351135153

Heriyanto, H. (2018). Thematic analysis sebagai metode menganalisa data untuk penelitian kualitatif. *Anuva: Jurnal Kajian Budaya, Perpustakaan, dan Informasi, 2*(3), 317–324. https://doi.org/10.14710/anuva.2.3.317-324

Huber, A. R., & Ward, Z. (n.d.). Non-consensual intimate image distribution: Nature, removal, and implications for the Online Safety Act. https://doi.org/10.1177/14773708241255821

Im, J., Schoenebeck, S., Iriarte, M., Grill, G., Wilkinson, D., Batool, A., Alharbi, R., Funwie, A., Gankhuu, T., Gilbert, E., & Naseem, M. (2022). Women's perspectives on harm and justice after online harassment. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2), 1–23. https://doi.org/10.1145/3555775

Kasita, I. D. (2022). Deepfake pornografi: Tren kekerasan gender berbasis online (KGBO) di era pandemi Covid-19. *Jurnal Wanita dan Keluarga, 3*(1), 16–26. https://doi.org/10.22146/jwk.5202

Kvale, S., & Brinkmann, S. (2009). *InterViews: Learning the craft of qualitative research interviewing* (2nd ed.). SAGE Publications.

Lakshané, R. (2024). *Non-consensual intimate imagery: An overview* (SSRN Scholarly Paper No. 4809632). Social Science Research Network. https://papers.ssrn.com/abstract=4809632

Lee, J.-Y. (2026). *Silenced by non-consensual deepfakes: Perceived threat and online self-censorship among female users: Mediating roles of perceived control, belief in a just world, and passive SNS use.*

Maras, M.-H., & Logie, K. (2024). Countering the complex, multifaceted nature of nude and sexually explicit deepfakes: An Augean task? *Crime Science, 13*(1), 31. https://doi.org/10.1186/s40163-024-00226-6

McGlynn, C., Rackley, E., & Houghton, R. (2017). Beyond 'revenge porn': The continuum of image-based sexual abuse. *Feminist Legal Studies, 25*(1), 25–46. https://doi.org/10.1007/s10691-017-9343-2

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1609406917733847. https://doi.org/10.1177/1609406917733847

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. SAGE Publications.

SpaceXAI — Creators of Grok, the AI Chatbot. (n.d.). *SpaceXAI*.

Third hemispheric report on the implementation of the Belém do Pará Convention: Prevention of violence against women in the Americas: Paths to follow. (2017). Organization of American States.

Umbach, R., Henry, N., Shelby, R., Stevens, G., & Gonzalez-Pons, K. (2025). *AI-generated image-based sexual abuse: Perpetration and consumption across three regions* (SSRN Scholarly Paper No. 5958708). Social Science Research Network. https://doi.org/10.2139/ssrn.5958708

United Nations (with UNDP). (1994). *Human development report 1994*. Oxford University Press.

Viola, M., & Voto, C. (2023). Designed to abuse? Deepfakes and the non-consensual diffusion of intimate images. *Synthese, 201*(1), 30. https://doi.org/10.1007/s11229-022-04012-2

© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution Share Alike (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/).