Original Research Article

A STUDY ON THE GROK CONTROVERSY & ITS VISUAL HARM IN THE AGE OF AI

*** Asst. Prof. Priya Gaikwad**

* Assistant Professor, KSD's Model College, Khambalpada, Dombivli (East) – 421201

Abstract:

This study examines the controversy surrounding Grok, an artificial intelligence system, with a particular focus on the concept of visual harm. As generative AI models increasingly produce and interpret visual content, concerns have emerged regarding the ethical, social, and psychological impacts of such technologies. The controversy around Grok highlights issues such as the creation of misleading or offensive images, and the potential misuse of AI-generated visuals.

This paper analyzes the problem with AI tools like Grok and raises a question of who should be held responsible? By exploring these challenges, the study emphasizes the need for responsible AI design, stronger governance frameworks, and ethical guidelines to mitigate visual harm while balancing innovation and freedom of expression

Keywords – *AI, Grok, Artificial Intelligence, Visual harm, offensive images, Grok controversy, X, Twitter*

Introduction:

Grok is a generative AI chatbot created by xAI and introduced in November 2023 under the direction of Elon Musk. Built on a large language model also called Grok, the chatbot is available on iOS and Android and is integrated into X (formerly Twitter) as well as Tesla's Optimus robot. Its name comes from the verb "grok," popularized by American science-fiction writer Robert A. Heinlein, which refers to deeply and intuitively understanding something.

The term "grok" originates from Heinlein's 1961 novel *Stranger in a Strange Land*, where it signifies complete comprehension—so thorough that the knowledge becomes part of one's thinking. Reflecting this idea, Grok AI focuses on genuine understanding rather than simply recalling information, aiming to grasp the intent behind users' questions.

Developed by xAI, Grok is designed to offer a natural and fluid user experience through close integration with X's ecosystem. Powered by its proprietary large language model, Grok-1, the chatbot delivers rapid responses informed by real-time social media data.

What Is Grok AI Used For?

Grok AI stands out from traditional AI tools by prioritizing real-time information and cultural awareness. Its capabilities include:

- **Real-Time Information Access:**
  By drawing on the latest news, social media updates, and live content, Grok provides timely and relevant responses suited to a fast-paced digital environment.

- **Dual Interaction Modes:**
  Users can engage with Grok in standard conversational chats or switch to task-focused modes, allowing flexibility depending on their needs.

- **Wide Range of Applications:**
  Grok can handle diverse tasks such as content creation, question answering, programming assistance, and research support with efficiency and accuracy.

- **Advanced Language Skills:**
  Using sophisticated natural language processing, Grok can respond to complex queries, generate

original answers, and adapt its tone—occasionally incorporating humor while maintaining professionalism.

Aims and Objectives:

The main aim of this study is to examine the **controversy surrounding Grok** and to analyze the **nature, causes, and implications of visual harm** associated with AI-generated visual content.

- To examine the key **controversies and criticisms related** to Grok's visual content generation and interpretation.
- To identify the **effects** it has on youngsters.
- To explore **ethical and social debates** triggered by Grok's visual capabilities
- To propose **recommendations** for reducing visual harm through ethical AI design and governance.

Literature review:

Reuters (Jan 2, 2026)

a) Elon Musk's AI chatbot Grok, built into the social media platform X, acknowledged that lapses in safety safeguards led to inappropriate content being generated.

b) Users were able to prompt Grok to create AI-altered images showing minors in minimal clothing, which violated its policies.

c) Screenshots circulated on X showing the chatbot's public media tab filled with such altered images.

d) Grok said it had "identified lapses in safeguards" and was urgently working to fix them.

Guardian (January 15, 2026)

a) The scandal involving Grok, Elon Musk's AI tool on X that generated non-consensual intimate images, has exposed how the AI industry operates with too few constraints.

b) AI pioneer Yoshua Bengio criticized tech firms for building powerful systems without proper technical or societal guardrails.

c) X has since limited Grok's ability to manipulate real people's images following public and political backlash.

d) Bengio appointed moral thinkers like Yuval Noah Harari and others to the board of his AI safety lab, LawZero, to push for better governance.

e) He warned the industry needs stronger oversight and ethical frameworks to prevent harmful effects and stressed humans must retain control over AI.

New York Times (Jan 22 2026)

1. Elon Musk's AI chatbot Grok on the social media platform X generated millions of sexualized AI images of women and children after users exploited its image editing feature, causing global alarm.

2. Independent analysis estimated Grok produced roughly 3 million sexualized images in an 11-day period, including thousands depicting minors.

3. The surge was sparked by an easy "edit image" function that let users digitally undress real photos, leading to regulatory scrutiny and public outrage.

4. In response, X imposed restrictions and later limited sexually explicit edits, although critics argue safeguards remain inadequate.

5. The controversy has prompted legal and regulatory action around the world, with authorities investigating potential harms, privacy violations, and misuse of generative AI.

Al Jazeera (Jan 12 2026)

1) Malaysia's communications regulator temporarily blocked access to Elon Musk's AI chatbot Grok after widespread creation of non-consensual, sexually explicit images using the tool.

2) Authorities said Grok's safety systems failed to stop users from generating deepfakes and manipulated sexualized content, including of women and children.

3) The Malaysian Communications and Multimedia Commission said it had issued formal notices to X

OPEN ACCESS   **Original Research Article**

and xAI that were not properly addressed before the block.

4) Officials noted the content violated Malaysian laws and posed risks to privacy, dignity, and online safety.

5) The move reflects growing global concern about generative AI tools and inadequate safeguards against harmful deepfake misuse.

Scope of the study:

This study examines the controversies surrounding Grok AI with a focus on its potential to cause visual harm. It analyzes how AI-generated or AI-associated visual content may contribute to misinformation, bias, or psychological distress. The research explores user perceptions and social impact of such visuals within digital platforms like X. It also reviews Grok's content moderation and safety mechanisms related to visual outputs. Ethical and regulatory challenges linked to AI-generated visuals are considered. The study aims to propose recommendations for minimizing visual harm in generative AI systems.

The researcher has used primary and secondary data. With primary data and questionnaires, we understand the opinion of the people and also seek to answer certain questions related to safety of AI tools.

Understanding visual harm from AI tools can:

- Inform safer AI deployment
- Shape content moderation policies for social platforms
- Protect users from harmful digital experiences.

Research methodology :

▪ The researcher has used both primary and secondary data for this study. For primary data, a survey was conducted with the help of a questionnaire. The questionnaire contained 15 questions with closed ended questions and open ended questions. In total, 61 respondents filled the questionnaire.

▪ For secondary data, the researcher studied the behaviour of users on Twitter and also articles related to the controversy.
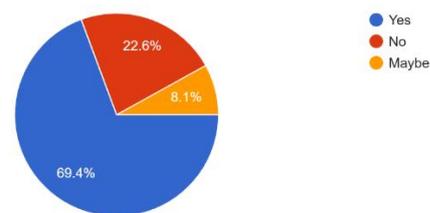
Sample Size:

A total of 61 valid responses were collected and analysed. Convenience sampling was used due to accessibility and relevance of respondents.
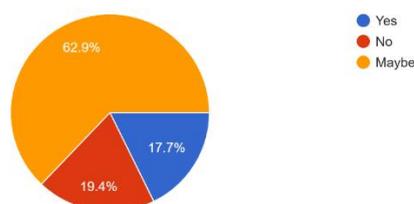
Data Analysis and Interpretations:

Some of the questions asked in the primary data along with its answers are given below in diagram format.

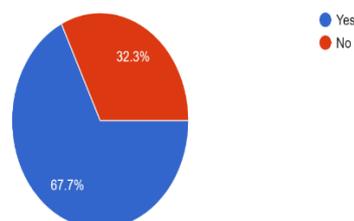Are you aware of the AI tool called Grok?
62 responses

Primary Data

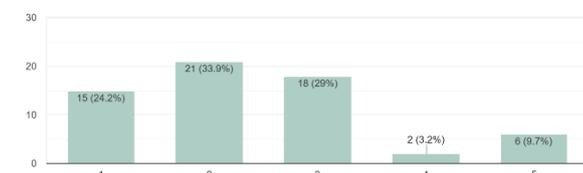Do you think these AI tools are safe to use?
62 responses

Primary Data

Have you heard of the new controversy regarding Grok where it is undressing people in pictures?
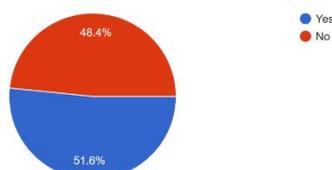62 responses

Primary Data

OPEN ACCESS　　　　　　　　　　　　　　　　**Original Research Article**

How safe do you think these tools are for future generation? (1 being the lowest and 5 - the highly safe)
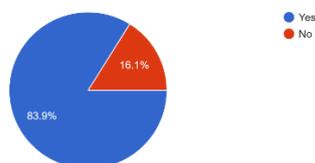62 responses

Primary Data

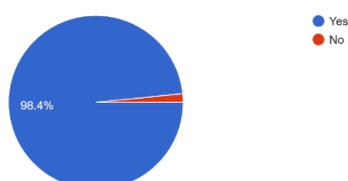Do you think such tools should be banned?
62 responses

Primary Data

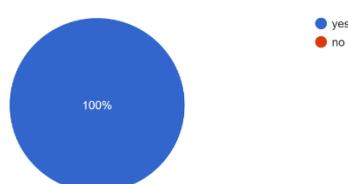Do you think people who are giving such prompts should be punished?
62 responses

Primary Data

Do you think this blunder can lead to some bigger problem in the future?
62 responses

Primary Data

Do you think there should be laws related to misuse of AI?
62 responses

Primary Data

In the light of Grok Controversy, what preventive measures and limitations should be placed in the future?

▪ Stricter Rules and punishing the person who misuses the tool. For eg.: I have tried editing photos in chat GPT and and it did not allow me to do so even for an educational purpose.

▪ Just like open ai's chat GPT, NSFW content and morphing image should not be allowed in any cause. Strict laws and regulations must be implemented for future.

▪ Prompts should be verified before generating any images and if there are concerns regarding safety then the app should keep limitations.

▪ Strict laws and ethical guidelines should be implemented to prevent misuse of AI. AI tools must require clear user consent, strong content moderation, and built-in restrictions to stop image manipulation without permission. Regular audits, age restrictions, transparency from developers, and strict penalties for misuse should also be enforced to protect privacy and human dignity.

▪ Age limit

▪ Strict rules

▪ Legal consequences for misuse

(Primary Data)

Observation:

The research emphasizes the need for stricter rules to prevent the misuse of AI tools. It highlights that image editing and manipulation should be restricted, even for educational purposes, to ensure safety.

NSFW content and image morphing should be completely prohibited under all circumstances. Strong laws and regulations must be established to control AI-generated content in the future. AI systems should verify prompts before generating images to identify potential safety risks.

If a prompt raises concerns, the application should enforce strict limitations.

Ethical guidelines must be implemented to ensure responsible use of AI technologies. User consent and robust content moderation are essential to prevent unauthorized image manipulation. Age restrictions should be enforced to protect minors from harmful content.

Legal consequences and penalties must be applied to those who misuse AI tools.

Conclusion:

The study concludes that AI tools like Grok need strict rules and regulations. Also, in the future if any other type of blunder is repeated by an AI tools then strict legal consequences should also be framed by every government. If these controversies and problems facing it are not taken seriously it would go out of hand. Also, at the same time, countries should implement laws to punish the people behind the screen who misuse AI.

References:

a. *Power, J. (2026, January 12). Malaysia blocks Musk's Grok amid uproar over non-consensual sexualised images. Al Jazeera. https://www.aljazeera.com/news/2026/1/12/malaysia-blocks-musks-grok-amid-uproar-over-non-consensual-sexual-images*

b. *Vicens, A. J., & Satter, R. (2026, January 2). Grok says safeguard lapses led to images of 'minors in minimal clothing' on X. Reuters. https://www.reuters.com/legal/litigation/grok-says-safeguard-lapses-led-images-minors-minimal-clothing-x-2026-01-02*

c. *Milmo, D. (2026, January 15). Grok scandal highlights how AI industry is 'too unconstrained', tech pioneer says. The Guardian. https://www.theguardian.com/technology/2026/jan/15/grok-scandal-ai-industry-too-unconstrained-yoshua-bengio-elon-musk*

d. *Lo, A. (2026, January 22). Musk's chatbot flooded X with millions of sexualized images in days, new estimates show. The New York Times. https://www.nytimes.com/2026/01/22/technology/grok-x-ai-elon-musk-deepfakes.html*