The Indian **EXPRESS**

JOURNALISM OF COURAGE

Grok, unhinged! Who is responsible for AI chatbot's sensational responses on X?

So, when Grok uses a misogynist Hindi expletive when responding to a user about their most prominent mutuals, or calls Musk one of the biggest sources of misinformation on the social media platform – it led to people asking Grok a flurry of questions, directly through their posts, or as comments to other posts.

Written by **Soumyarendra Barik** (Follow)

New Delhi | Updated: March 22, 2025 07:06 IST

✅**NewsGuard**

🕐 5 min read

f   X   ✆   ⬮

3/22/25, 8:18 AM

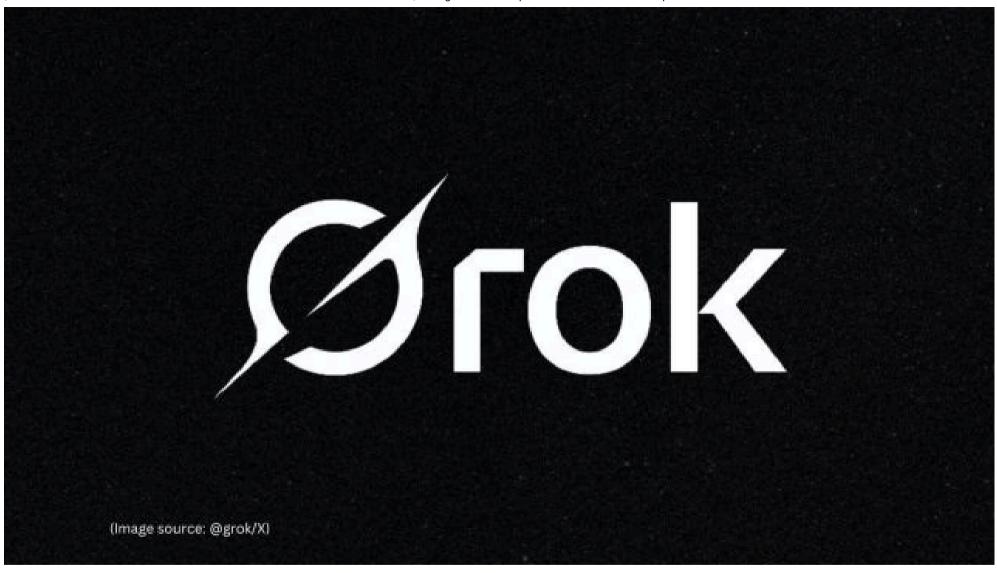Grok, unhinged! Who is responsible for its sensational responses on X?

Grok is Elon Musk's AI venture (Image source: @grok/X)

With the Indian government in touch with Elon Musk's X over stirring responses generated by its artificial intelligence (AI) chatbot Grok, the question that many in the government are grappling with is: Who is actually responsible for the retorts that the AI has been producing on the social media platform?

Laden with profanities, sweeping takes, colouring some conservative users, including X founder Musk, as the biggest spreaders of misinformation — Grok's responses to questions posed by Indian users have so far turned out to be an amalgamation of the attitudes and demeanour familiar to those who frequently use the social network.

Here are some facts: Grok is not a person, at best, it is a computer code running on high-end compute at the back end, and at its worst, that code periodically churns out the underbelly of the data that has been fed to it. Grok is artificial, its intelligence debatable.

So, when Grok uses a misogynist Hindi expletive when responding to a user about their most prominent mutuals, or calls Musk one of the biggest sources of misinformation on the social media platform — it led to people asking Grok a flurry of questions, directly through their posts, or as comments to other posts.

This piece aims to demystify three main concerns around what's happening with Grok: who is responsible for its responses, are the people asking it questions somehow liable, and if Grok is a source of truth.

Who is liable, can people be penalised?

Internet platforms such as X, Meta, and YouTube have legal protection from the content that their users post. This, in law, is called safe harbour — the argument being platforms have no control over what users are posting. They are mere conduits, so they can not be held liable for hosting third-party content.

While that convention itself is currently being debated given virality and the potential of speech on such platforms to cause real world harms, the million dollar question is whether Grok, an artificial output generator, can have safe harbour protections.

That is a complex question to deal with for lawmakers. X has told the Indian government that it has been trained on the open-Internet, which presumably also includes content that users post on X. So, in a way, everything that Grok generates is based on

what people spending years on the Internet have produced. But can then they be held responsible? That is like asking if the ocean can be sued for being wet.

Besides, speech is a highly protected category in India, with the Indian Constitution affording the freedom of expression as a fundamental right, with some reasonable restrictions. But, those rights are available to human beings. Humans' speech should be censored only under select circumstances when they obviously violate restrictions laid down in the Constitution. Does Grok have the right to unfettered free speech? And what is Grok's free speech even? Its code essentially determines what the next word in a sentence should be, which is a factor of the underlying dataset it has been trained on, which in turn is generated by actual humans. Both the code, and the content in the language model.

So, many would argue that the liability of Grok's responses primarily lies with xAI, its creators, and X for allowing Grok to produce responses without any filters. But that too raises some pertinent questions. How does one hold creators of an algorithm responsible? Is it the highly-paid people who have written the code, or the low-wage data annotators? These are questions that regulators around the world are unlikely to have a quick, and accurate answer to. "Grok is certainly not a real person, it's an artificial entity. But some of its responses are definitely problematic. It's an interesting, and difficult problem, that us in government will have to figure out," a senior government official said.

Should you trust Grok?

The short answer to that question is, AI responses should not be treated as accurate pieces of information, no matter how much they satiate one's socio-political beliefs. Already, platforms are applying filters on their AI models to restrict their political speech in order to stay safe from government scrutiny.

As India headed to Lok Sabha elections last year, [Google](#) said it will restrict the types of election-related questions users can ask its artificial intelligence (AI) chatbot [Gemini](#) in the country. Earlier, Krutrim, the chatbot developed by an Indian AI startup founded by Bhavish Aggarwal of Ola, had been found to self-censor on certain keywords. AI platforms are built to predict the next

3/22/25, 8:18 AM

Grok, unhinged! Who is responsible for its sensational responses on X?

word in a phrase, and to try and satisfy the query a user has asked — and models like Grok have so far shown they will do anything to achieve that, beg, borrow or steal.

Soumyarendra Barik (Follow)

𝕏

Soumyarendra Barik is Special Correspondent with The Indian Express and reports on the intersection of technology, policy and society. With over five years of newsroom experience, he has reported on issues of gig workers' rights, privacy, India's prevalent digital divide and a range of other **... Read More**

© The Indian Express Pvt Ltd

This article went live on March twenty-first, twenty twenty-five, at thirty-six minutes past twelve in the afternoon.

TAGS: **Artificial Intelligence**   **Express Premium**