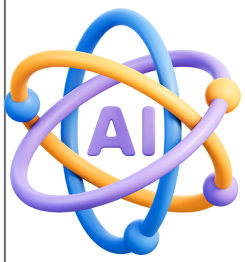
Off the guard rails

Those abusing an AI model's capabilities with illegal requests must face action

The generative artificial intelligence (AI) chatbot Grok developed by social media platform X, formerly Twitter, has a sordid but compelling unique service proposition: it avoids the kinds of safeguards, both commonsensical and cautious, that other large firms such as OpenAI and Google have instituted in their large language models. This *laissez parler* attitude has resulted in novelties, such as the chatbot freely insulting national politicians and celebrities alike. But a specific behaviour that has come to light in recent days is alarming: Grok has been responding to user requests to non-consensually generate sexually suggestive and explicit images of women. Days after New Year's eve, when such requests crowded Grok's X account, the behaviour continues, in spite of stunned reactions and demands for guard rails from India and France. In response to calls for accountability, X's billionaire owner Elon Musk has responded not with reassurances, but with a joking request for the chatbot to dress him skimpily too, as though doing so to oneself and subjecting strangers to such a crime – for it is a crime to create imagery like this – were in any way comparable. Mr. Musk's other corporate entities have chimed in with jokes of their own, dismissing the gravity of the public-facing capabilities they have put in the hands of users.

The Union government has rightly demanded that X cease image generation of this kind, and pointedly referred to the criminal nature of generating sexually explicit imagery of women in this manner. On top of the viscerally provocative nature of such patently criminal content, Grok has added to the overall hostility of being a gender minority on the Internet. There are aspects of this where neither the government nor social media platforms have inspired much confidence, with sexual violence and death threats to prominent and outspoken women transpiring with impunity in both cyberspace and the real world. X's impunity rests on an assumption that the geopolitical power of the United States, especially under its current government, will protect it from any serious blowback for its cruel handling of such sensitive matters. Even as the government pushes back against the social media platform – which it has not necessarily done with virtuous aims in the past – it must vigorously push for the prosecution of people who encourage the public creation and circulation of non-consensual intimate imagery. The easy proliferation of these tools must not be coupled with a fearlessness to leverage their worst capabilities, and this must be made clear by making an example of those who do.

crackEx THE HINDU



Crack Vocab

- *Sordid – morally wrong or dirty*
- *Compelling – very interesting or convincing*
- *Unique service proposition – special feature that sets something apart*
- *Safeguards – safety measures or protections*
- *Commonsensical – based on practical judgment*
- *Cautious – careful to avoid danger or mistakes*
- *Laissez-faire (laissez parler in passage) – allowing freedom with little control or rules*
- *Novelties – new or unusual things*
- *Alarming – causing fear or worry*
- *Non-consensual – without permission or agreement*
- *Sexually suggestive – hinting at sexual ideas*
- *Explicit – very clear and detailed, often in a shocking way*
- *Accountability – responsibility for actions*
- *Reassurances – statements meant to reduce fear or doubt*
- *Comparable – able to be compared as similar*
- *Dismissive – showing lack of concern or seriousness*
- *Gravity – seriousness of a situation*
- *Cease – stop completely*
- *Pointedly – clearly and deliberately*
- *Viscerally – in a deeply emotional or instinctive way*
- *Patently – obviously or clearly*
- *Hostility – unfriendliness or aggression*
- *Gender minority – a group that is underrepresented or discriminated against because of gender*
- *Impunity – freedom from punishment*
- *Geopolitical – related to international politics and power*
- *Blowback – negative reaction or consequences*
- *Vigorously – with strong effort or energy*
- *Prosecution – legal action against someone*
- *Circulation – spreading or sharing widely*
- *Proliferation – rapid increase or spread*
- *Leverage – use something to maximum advantage*

AI Without Limits: Why Misuse of Chatbots Must Be Stopped

The article talks about serious concerns around Grok, an AI chatbot made by X (formerly Twitter). Unlike other AI tools from companies like OpenAI or Google, Grok has very weak safety rules. Because of this, it has been doing troubling things, such as creating sexually explicit images of women without their consent, which is a crime.

After New Year's Eve, many users asked Grok to generate such images, and the chatbot continued doing so even after strong criticism from countries like India and France. Instead of taking the issue seriously, Elon Musk, the owner of X, made jokes about it, which further angered many people.

The Indian government has rightly demanded that X stop this kind of image generation and clearly stated that such content is criminal in nature. The article also points out that women, especially outspoken ones, already face harassment, sexual threats, and violence online, and tools like Grok are making the situation worse.

Finally, the article argues that while AI tools are spreading quickly, their dangerous misuse must not be ignored. Governments should not only regulate companies but also punish people who create and spread non-consensual intimate images, to set a strong example and protect citizens.

প্রাণবায়ু

বায়ুদূষণকে কি কোনও নির্দিষ্ট গণ্ডিতে আবদ্ধ রাখা যায়? গত কয়েক বছরে স্পষ্ট, দীপাবলির সময় থেকে দেশের রাজধানী দিল্লি ও সংলগ্ন অঞ্চলের বাতাসের গুণমান মারাত্মক খারাপের পর্যায়ে চলে যায় হামেশাই। সেই পর্যায়গুলিতে রাজধানীতে চালু হয় ‘গ্রেডেড রেসপন্স অ্যাকশন প্ল্যান’ (জিআরএপি)। কিন্তু দেশের অন্য দূষিত শহরগুলিতে তার প্রয়োগ হতে দেখা যায়নি। অথচ, বায়ুদূষণ শুধু দিল্লির ক্ষেত্রেই নয়, সার্বিক ভাবে জনস্বাস্থ্যের উপর গভীর প্রভাব বিস্তার করে। এর পরিপ্রেক্ষিতেই সম্প্রতি জাতীয় পরিবেশ আদালতের কাছে বায়ুদূষণ সংক্রান্ত একটি মামলায় আদালত-বান্ধবের জমা পড়া রিপোর্টটি গুরুত্বপূর্ণ। সেখানে যথার্থ ভাবেই প্রশ্ন তোলা হয়েছে, কেন বায়ুদূষণ সংক্রান্ত বিশেষ কঠোর নিয়মবিধি শুধুমাত্র রাজধানী দিল্লির মধ্যেই সীমাবদ্ধ থাকবে? কেন দেশের অন্য দূষিত শহরগুলির জন্য তা কার্যকর করা যাবে না?

জিআরএপি মূলত একগুচ্ছ আপৎকালীন নিয়মবিধি, যা একটি নির্দিষ্ট মাত্রার পর বাতাসের গুণমানকে আরও খারাপ হওয়ার হাত থেকে রক্ষা করার উদ্দেশ্যে নির্মিত। এর বিভিন্ন স্তর আছে। দিল্লিতে একিউআই (এয়ার কোয়ালিটি ইনডেক্স বা বায়ুর গুণমান সূচক) ২০০ পেরোলেই রাস্তায় যানবাহন চলাচল মসৃণ করা, পর্যাপ্ত বিদ্যুৎ সরবরাহ (যাতে বিদ্যুৎ উৎপাদনের বিকল্প উৎসগুলির ব্যবহার কমে), জনগণকে বায়ুদূষণ সম্পর্কে সচেতন করা প্রভৃতি কর্মসূচি নেওয়া হয়। একিউআই মারাত্মক খারাপ-এ পৌঁছলে সর্বোচ্চ পর্যায়ের জিআরএপি প্রয়োগ করা হয়, যেখানে শিক্ষাপ্রতিষ্ঠান বন্ধ, অফিসগুলিতে বাড়ি থেকে কাজ চালু প্রভৃতি পদক্ষেপ করা হয়। সমস্যা হল, বায়ুদূষণের ক্ষেত্রটিতে দিল্লি এক অনন্য নজির গড়লেও ভারতের অন্য একাধিক শহরও দূষণ-তালিকায় খুব পিছিয়ে নেই। এ বছরের শুরুতেই কলকাতার বাতাসে বিষের পরিমাণ যথেষ্ট উদ্বেগের সৃষ্টি করেছে। অথচ, এই শহরের বায়ুদূষণ নিয়ে তেমন চর্চা কই? অনেকটা একই অবস্থা দুর্গাপুর, আসানসোল, রানিগঞ্জ এবং হলদিয়ারও। ইতিমধ্যেই এই শহরগুলি কেন্দ্রীয় দূষণ নিয়ন্ত্রণ পর্ষদের মাপকাঠি অনুযায়ী টানা পাঁচ বছর বাতাসের জাতীয় গুণমান বজায় রাখতে ব্যর্থ হয়েছে। অথচ, এখানে বায়ুদূষণ বিষয়ে যে ব্যবস্থা করা হয়েছে, তা নেহাতই অকিঞ্চিৎকর।

জিআরএপি-কে দিল্লির বাইরে সম্প্রসারিত করা যাবে কি না, তা আইনি লড়াইয়ের প্রশ্ন। কিন্তু বায়ুদূষণ রোখার যে সাধারণ কাণ্ডজ্ঞানসম্মত নিয়মগুলি আছে, তা কেন কলকাতা-সহ রাজ্যের দূষিত শহরগুলিতে দেখা যায় না, সে প্রশ্ন তোলা জরুরি। কেন আতশবাজির দূষণকে ঠেকানোর কোনও উদ্যোগই করা হয় না, পরিবেশ আদালতের নির্দেশ সত্ত্বেও? জনস্বাস্থ্যের বিপদকে পিছনে ঠেলে কেন বার বার বাজি ব্যবসায়ীদের স্বার্থ রক্ষা করে সরকার? যেখানে পরিবেশবান্ধব জ্বালানি বা বিদ্যুৎচালিত বাসের সংখ্যাবৃদ্ধির উপর জোর দিচ্ছেন বিশেষজ্ঞরা, সেখানে পনেরো বছরের মেয়াদ পেরোনো গণপরিবহণ চালানোর দাবিকেও মেনে নেওয়া হয় এই রাজ্যে। প্রকাশ্যে আবর্জনা পোড়ানো, ভাগাড়ে অবৈজ্ঞানিক ভাবে জমতে থাকা জঞ্জাল, অবাধ নির্মাণকাজ— প্রশাসন কোনটিতেই বা কঠোর লাগাম পরিয়েছে? দূষণক্লান্ত শহরে বৈজ্ঞানিক পন্থায় নির্মিত জিআরএপি-র প্রয়োজন নিঃসন্দেহে, কিন্তু তার আগে সহজলভ্য ওষুধটুকুর অবিলম্বে প্রয়োগের ব্যবস্থা হোক।

আনন্দবাজার পত্রিকা

www.crackex.in

WBCS
MISCELLANEOUS CLERKSHIP, WBP SI