

AI Causing Chaos

“ So much AI turns out to be low-waged people in a call center in the Global South pretending to be robots that Indian techies have a joke about it: “AI stands for ‘absent Indian’” - [Cory Doctorow](#)

I am collecting examples of AI fails by people who either deliberately or naively mislead about the capabilities of AI.

2024

- **11/02/24** - [System set up to detect cheating students found 97% of students were cheating and UK officials just accepted this](#)
- **09/02/24** - [Gemini LLM conflates unsafe in the C# memory management context with unsafe as in unethical.](#)
- **05/02/24** - [Using neural networks to generate images of fake ids for online verification](#)
- **04/02/24** - via [Yoav Goldberg](#) - [GEM using models to predict protected characteristics of job applicants](#) when such information is not supplied
- **28/01/24** [Dudesy Podcast](#) claimed that they trained an AI to impersonate the late comedian [George Carlin](#) but [Cory Doctorow](#) and [Simon Willison](#) both call bullshit. IT stinks of the "We made an AI watch 100 hours of X..." meme that went round a few years ago.
- **20/1/24** [DPD](#) introduce a ChatGPT-powered bot which users jailbreak and get to swear and call DPD 'useless'

2023

- **20/12/23** [Chevrolet integrated a ChatGPT-powered bot into their site](#) but because there is no such thing as input sanitization for [LLMS](#) people used prompt injection attacks to make the bot agree to sell users cars for \$1.
- **24/10/23** [Cruise robo-taxi dragged a person 20ft along the floor](#) and had their license to operate in San Francisco suspended.

- **17/01/23** it emerged that [a 2016 demo of Tesla's self-driving capabilities was completely staged](#)

-

Revision #4

Created 8 February 2024 20:59:33 by James

Updated 11 February 2024 22:54:25 by James