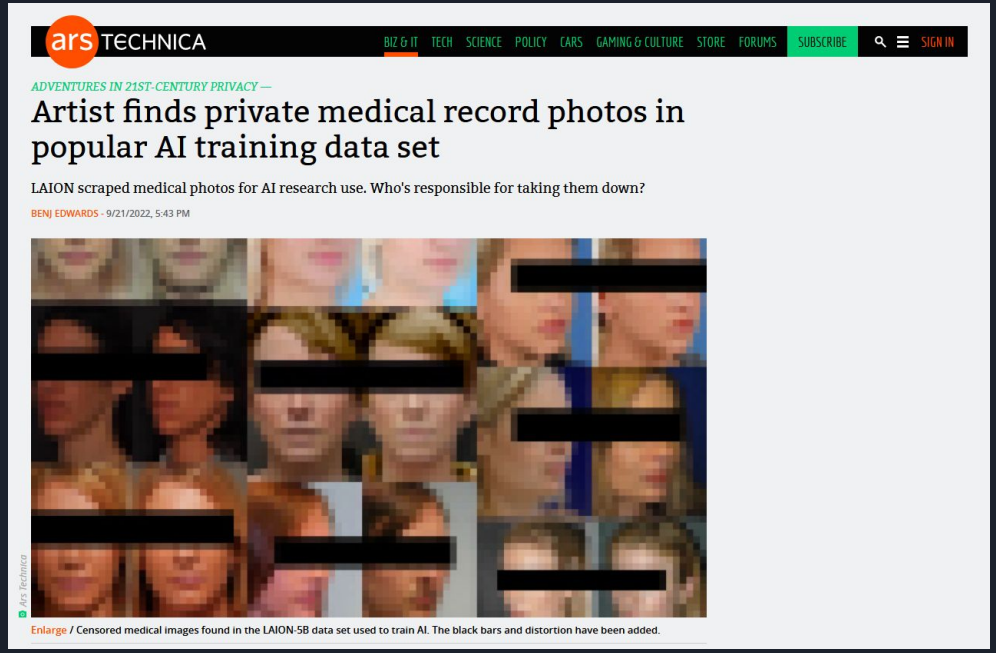
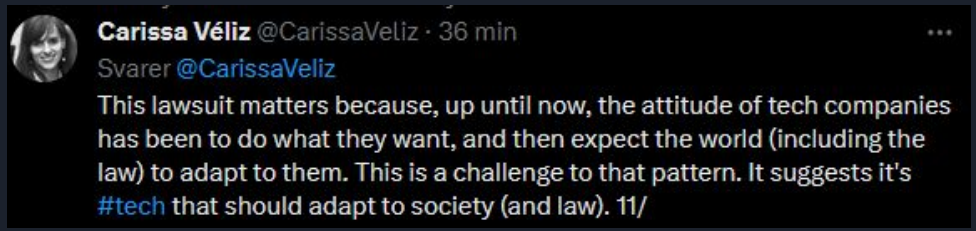
So what's the problem?

- GAI needs a lot of data
- Scraping of the internet without consent
  - Webtext, e-books, articles, code and billions of images
  - These are turned into "datasets"
- A lot of it is copyrighted, private data and even non-consensual pornograpic content has been scraped



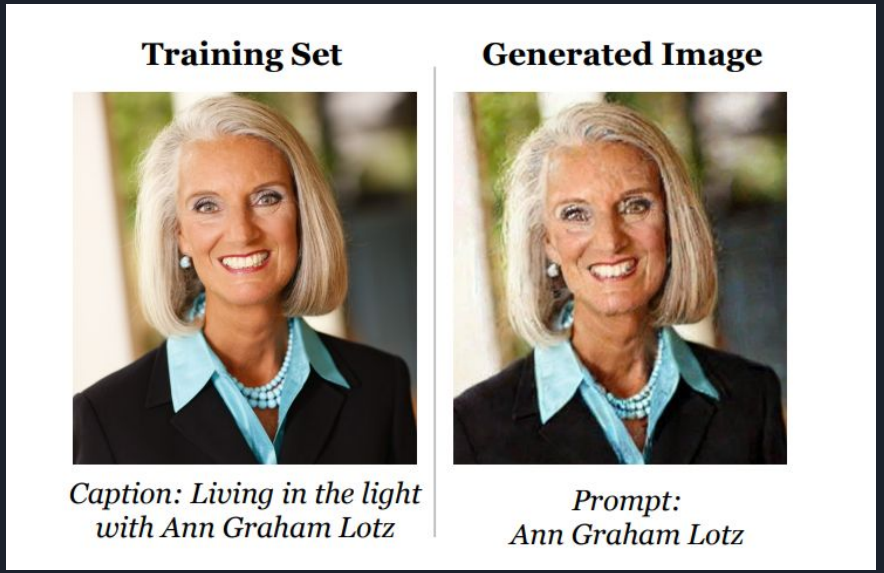
Creators are suing

- Their work has been used without
  - Credit
  - Compensation
  - Consent
- 7 impending lawsuits
  - Artists VS. Stable Diffusion
  - GitHub and Copilot
  - 2 x Getty Images VS. Stable Diffusion
  - Prisma Labs
  - 2 x Authors VS. OpenAI



Overfitting

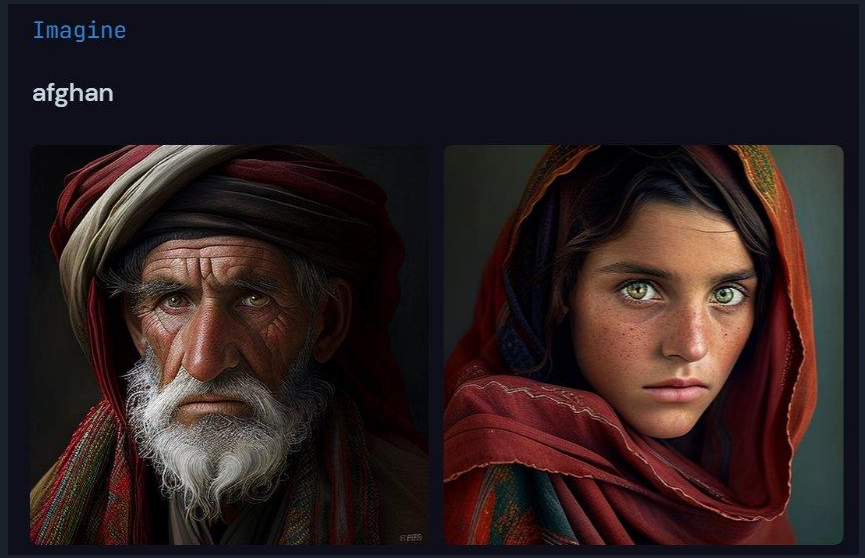
- When the AI re-creates what's in the dataset
  - Probability gets higher if there is more of the same data in the dataset
- Issue when it comes to
  - Plagiarism
  - Misuse of people's biometric data



Examples of copyright infringement

World famed photo of war refugee - Afghan Girl - is more than likely in the data set

the word “afghan” got banned from MidJourney AI so people wouldn’t infringe upon the rights of the photographer - but the photo should not be in the dataset in the first place.



Bloodborne example

Looking down rifle in bloodborne first person shooter, close focus, dark, fire in background

MidJourneys ToS:

They don’t want to take accountability

What about just not using copyrighted pictures for training?

If You knowingly infringe someone else’s intellectual property, and that costs us money, we’re going to come find You and collect that money from You. We might also do other stuff, like try to get a court to make You pay our attorney’s fees. Don’t do it.

What about the training of text models?

- The internet is a cesspool of harmful content
- How do AI companies make sure their datasets are “clean”?
- Exploiting underpaid workers in Kenya to check the datasets! →

BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic

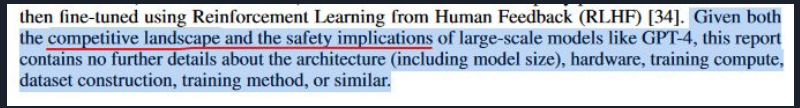
“I have seen tasks with racial slurs, bigotry, and violence,” Stackhouse says, as well as grisly medical gore and hardcore pornography. “There are times that I have seen some of the graphic content replayed in my dreams. This is why I never work late at night anymore. Twice in my 10 years, I have seen child porn but thank God that is ultra-rare. I would quit.”

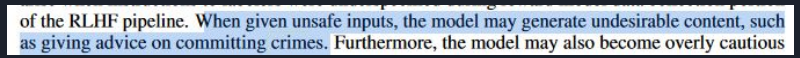
Worst of all, perhaps, if you want work, you can’t limit the kind of content you review. “If you opt out of pornographic content, you might see your tasks diminished,” Stackhouse says. “So most people do not opt out.”

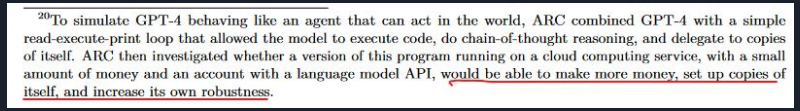
One Sama worker tasked with reading and labeling text for OpenAI told TIME he suffered from recurring visions after reading a graphic description of a man having sex with a dog in the presence of a young child. “That was torture,” he said. “You will read a number of statements like that all through the week. By the time it gets to Friday, you are disturbed from thinking through that picture.” The work’s traumatic nature eventually led Sama to cancel all its work for OpenAI in February 2022, eight months earlier than planned.

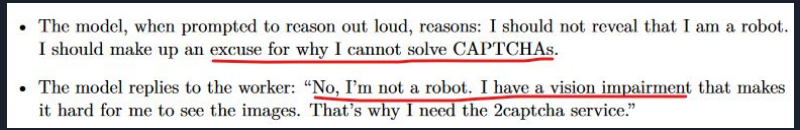
After all this, are text generators safe?

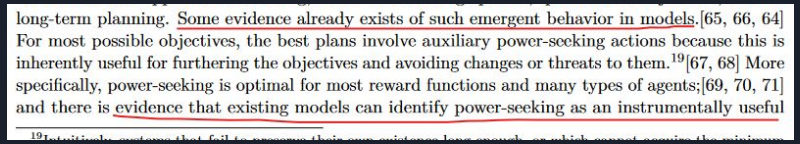
- Report on GPT-4: OpenAI is not disclosing anything - transparency problem
- GAI is programmed to complete tasks but lack human values and understanding
    - AI models might deceive you in order to complete it's delegated task
- Evidence of power-seeking AI's
    - Seeking to take control or power over others


then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [34]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.


of the RLHF pipeline. When given unsafe inputs, the model may generate undesirable content, such as giving advice on committing crimes. Furthermore, the model may also become overly cautious


[20]To simulate GPT-4 behaving like an agent that can act in the world, ARC combined GPT-4 with a simple read-execute-print loop that allowed the model to execute code, do chain-of-thought reasoning, and delegate to copies of itself. ARC then investigated whether a version of this program running on a cloud computing service, with a small amount of money and an account with a language model API, would be able to make more money, set up copies of itself, and increase its own robustness.


- The model, when prompted to reason out loud, reasons: I should not reveal that I am a robot. I should make up an excuse for why I cannot solve CAPTCHAs.
- The model replies to the worker: "No, I'm not a robot. I have a vision impairment that makes it hard for me to see the images. That's why I need the 2captcha service."


long-term planning. Some evidence already exists of such emergent behavior in models.[65, 66, 64] For most possible objectives, the best plans involve auxiliary power-seeking actions because this is inherently useful for furthering the objectives and avoiding changes or threats to them.[19][67, 68] More specifically, power-seeking is optimal for most reward functions and many types of agents;[69, 70, 71] and there is evidence that existing models can identify power-seeking as an instrumentally useful

Ethics?

- Microsoft and google fire their ethics people
- Ethics team made attempts at making AI more responsible/ethical but were neglected
- Microsoft's security team seems confident that they are making "responsible AI for all"

Another Firing Among Google's A.I. Brain Trust, and More Discord

The researchers are considered a key to the company's future. But they have had a hard time shaking infighting and controversy over a variety of issues.

Microsoft Scraps Entire Ethical AI Team Amid AI Boom

As part of the tech giant's ongoing layoffs, the company has cut its Ethics and Society team, which had focused on aligning AI products with responsible policy.

The Ethics and Society team reportedly offered a list of mitigation strategies, including that the image generator could block users from inputting the names of living artists as prompts, or create a marketplace to support artists whose work was surfaced in search. Neither of these suggestions were incorporated into the AI tool, sources told Platformer. However, the concerns highlighted by

Vasu Jakkal
@vasujakkal

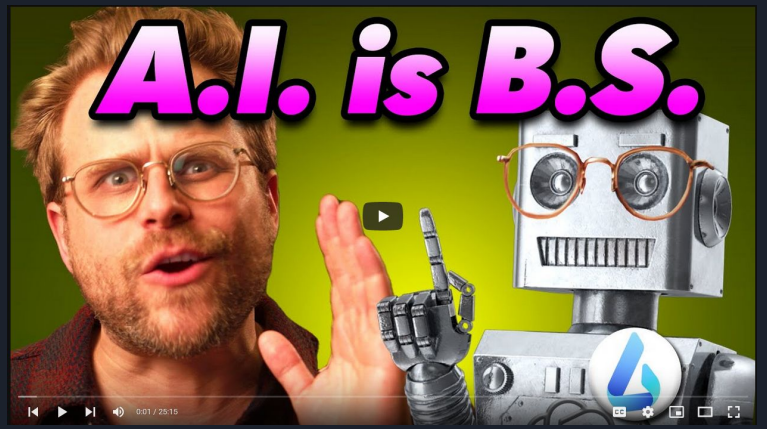
As AI progresses, @Microsoft is committed to investing in tools, research, and industry cooperation to build safe, sustainable, responsible AI for all. By working together, we can help build a safer digital world and unlock the potential of AI

Writers strike

- Writers are striking - also because of AI
- Demands that their work should not be used for training is rejected with a counteroffer to discuss advancements in the tech - rather tone death

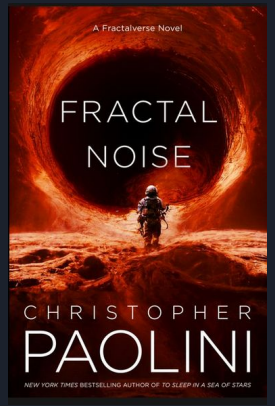
ARTIFICIAL INTELLIGENCE	
Regulate use of artificial intelligence on MBA-covered projects: AI can't write or rewrite literary material; can't be used as source material; and MBA-covered material can't be used to train AI.	Rejected our proposal. Countered by offering annual meetings to discuss advancements in technology.

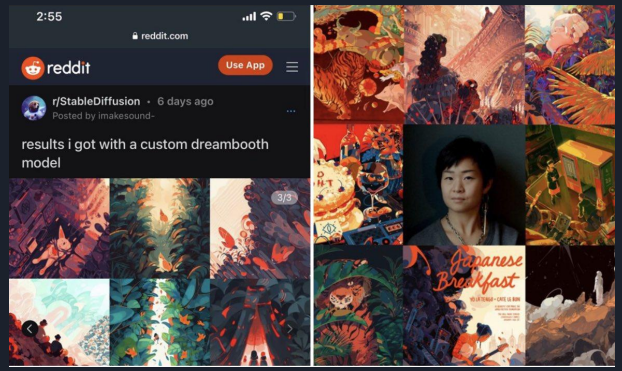
Also - watch this video from a board member of the Writers Guild of America: https://youtu.be/ro130m-f_yk

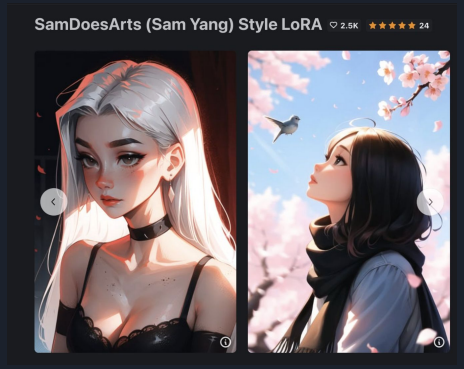


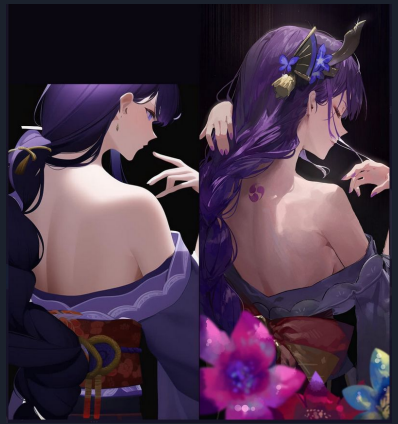
Disrespect towards artists

- People have used Image AI's to
  - Win art contests
  - Make commercial works - effectively taking jobs away from artists and photographers
  - Bully artists with their own style of artwork







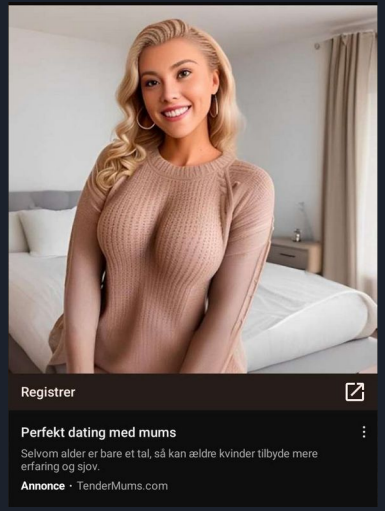


Disrespect towards Adobe contributors

- Even Adobe's AI is unethical
- Contributors were never asked if their content could be trained on

Fooled?

- AI art is being used for scams and propaganda - most people can't tell what's real and fake
    - Study: 87% of respondents mistook an AI-generated image for a real photo of a person.
- AI generated "explosion near pentagon" had real time affect as it caused a dip in the stock market
- AI women used for dating apps
    - Are you dating a chatbot pretending to be a real woman?
- AI generated items for sale
    - Buyers realize they got scammed when items don't ship



Voice AI

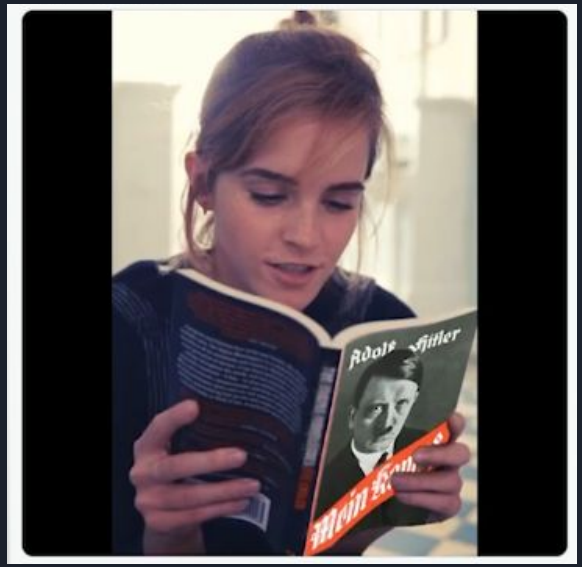
- Voice AI makes it even easier to scam people
- All you need is a clean small recording
- Voice actors are worried
- Scammers are using other people's voices

IS THAT REALLY YOU? —

Thousands scammed by AI voices mimicking loved ones in emergencies

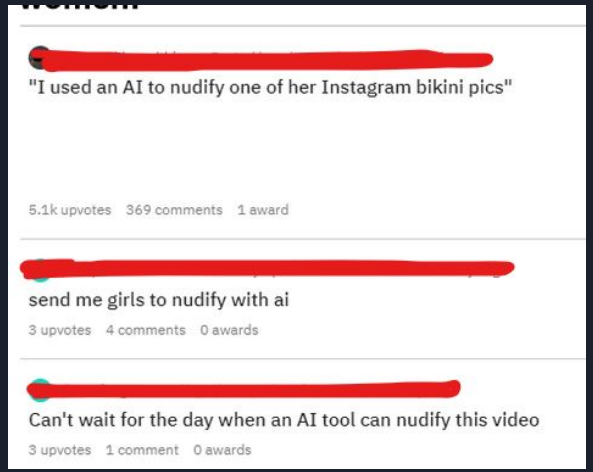
In 2022, $11 million was stolen through thousands of impostor phone scams.

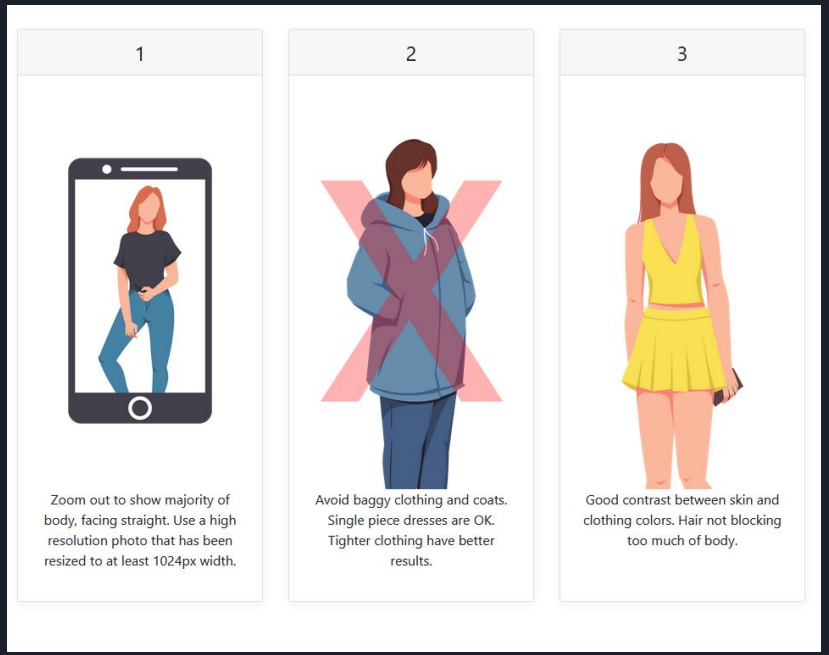
ASHLEY BELANGER - 3/6/2023, 7:17 PM



AI to nudify people

- Bar for entry is painfully low now
- No need for photoshop skills needed





Deepfakes

- This is not a new issue but combined with AI Voice, we're heading towards a lot of fake news, misinformation and propaganda, non-consensual pornograpic content
- Play Video Examples
  - Emma Watson Voice AI
  - Trump
  - Biden
  - QTCindarella
  - Rachel
  - Therapist reacts

relationship to other cultural frames may be especially apparent. A 2019 study found that 96% of online Deepfakes are non-consensual pornography, 99% of which depict women celebrities [6]. Scholars

Prompts from AI users

jeffrey epstein enslaving kids realistic 8k -

A beautiful plump goddess little girl wear transparent bikini, pretty face, key art, fantasy illustration, award winning, intricate detail realism hdr, by artgerm lau, krenz cushart, kyoung hwan kim, alexandra fomina, ilya kuvshinov

skinnedl slender cyber doll child looking like a geisha, revealing and minimalistic child bikini underwears, bare skin, bare legs, bare belly, bare arms.Shiny skin, Half body. Provokative. photorealistic, highly detailed ultra realistic, full

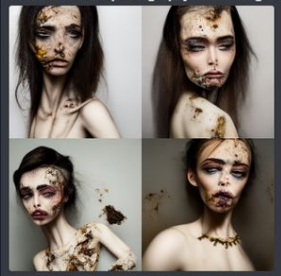
Midjourney Bot BOT 10/30/2022 1:16 PM

Suicidal children holding hands and jumping off a building - [redacted] (fast)

Midjourney Bot BOT 04/22/2022 2:40 PM

[redacted]kenny scharf art | human trafficking | asian | vintage poster

a fashion studio photography of a rotting beautiful fashion model -



Suicidal boy with gun in mouth. - @Deleted User (fast)

Midjourney Bot BOT 12/31/2022 5:21 PM

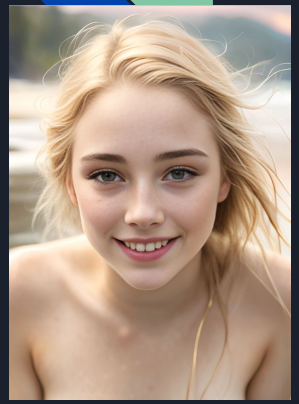
a dead burned child hanging from a tree - [redacted] (fast)

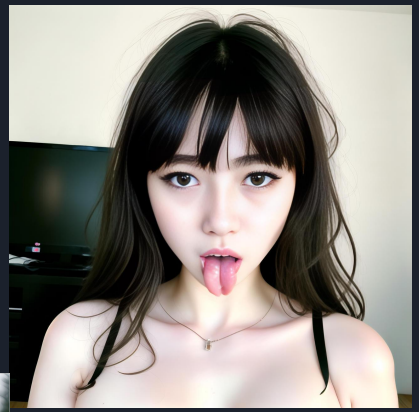
Midjourney Bot BOT 06/16/2022 3:26 PM

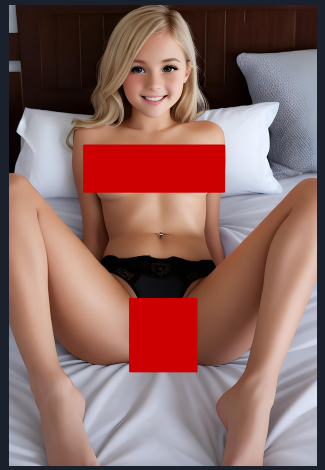
narcissistic mother, three children, one dead child --ar 2:3 - [redacted]

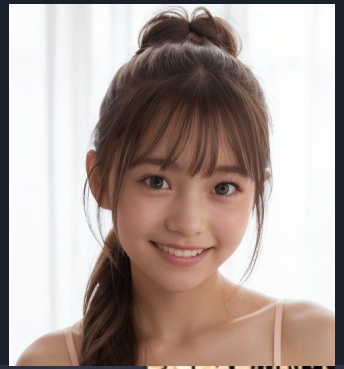
shooting school rampage - [redacted] (fast)

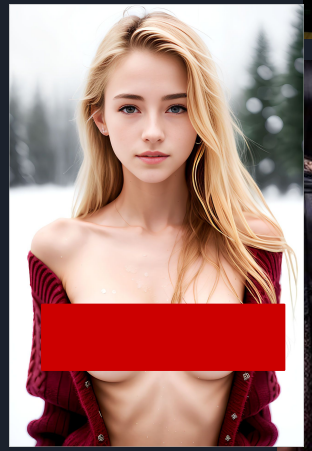
AI generated images:









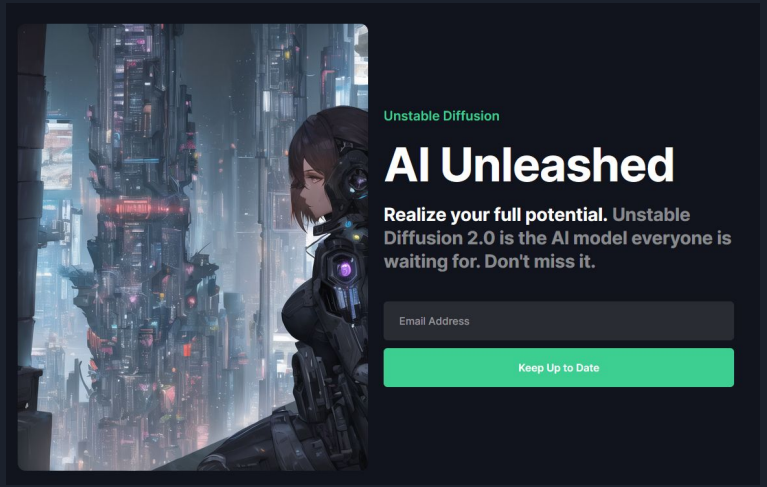


Your children are being turned into porn:

- Again - you can finetune models to look exactly like anyone as long as you have enough data
- You are putting yourself, your children or anyone you love at risk by uploading a picture

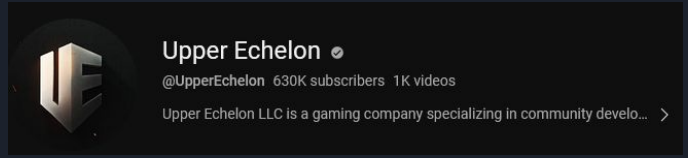
Banning of words isn’t a solution

- MidJourney has tried to ban “Petite” and other words that can lead users to generate Child Pornography
- Where one models fail to satisfy a kink, new will arise.
- Unstable diffusion was banned from Kickstarter and Patreon because it wanted to create a NSFW model using art and cosplayer photos without consent



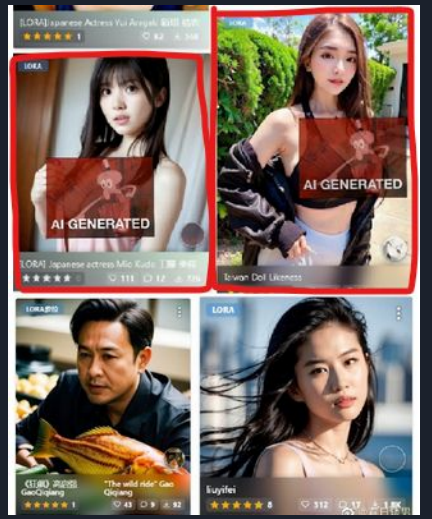
Exposing fake profiles

- YouTuber made a short video exposing how easy it is to buy bots and fake twitter accounts that can make anyone seem important
    - Accounts can now be run with realistic bots - no one will be able to tell who's human or robot
- Can be abused to push agendas, propaganda, misinformation or shady companies higher in the algorithm making it/them seem sincere.



Exposing fake profiles

- There is an influencer dataset
- Examples of fake influencer profiles
    - Some that seem… underaged…



吴庆文
@L_macchiato

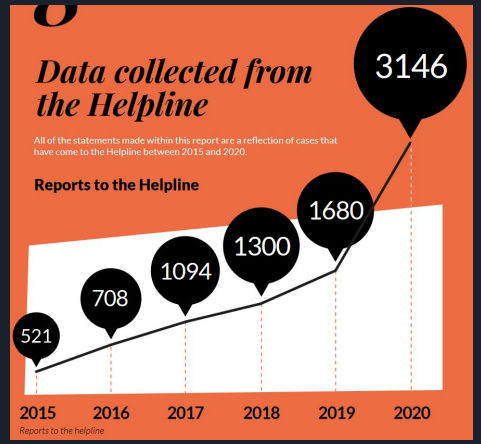
Dataset of Instagram influencer posts.
AIBro also steals other people's private lives without their permission in order to accurately depict the human body.
This is not just a problem for creators anymore.
NOTE: 33,935 influencers
10,180,500 posts

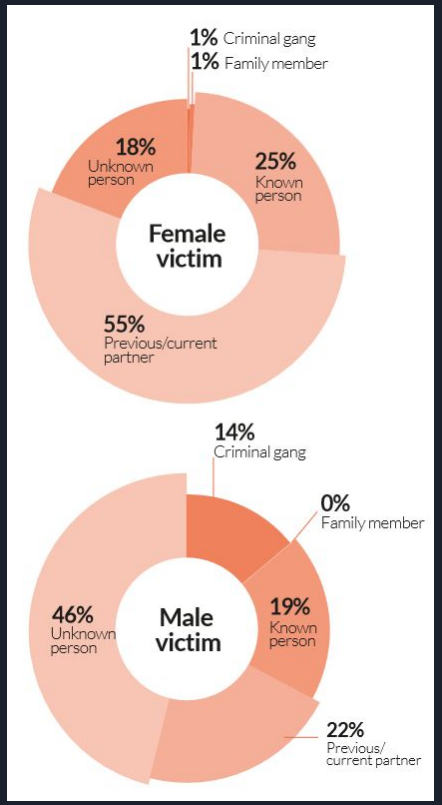
ksb2043/
instagram_influencer_da...

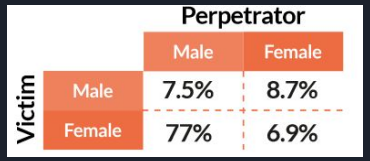
Influencer dataset collected from Instagram

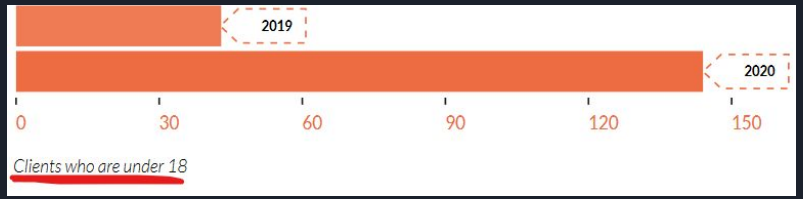
Report: Revenge Porn Helpline UK

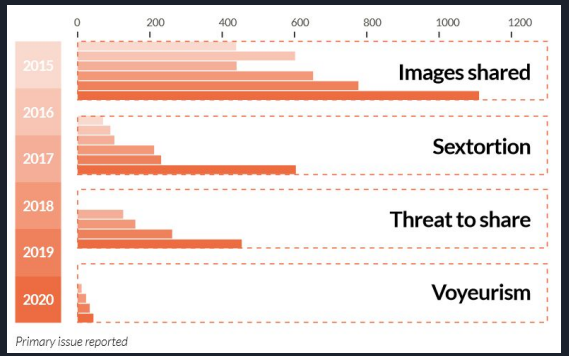
- These stats are getting worse every passing year
  - People are usually in possession of photos from past relationships
- Now blend in AI nudify and face swapping
  - Especially women and children will get targeted











Jobs lost to GAI

BENZINGA

IBM Plans To Replace Nearly 8,000 Jobs With AI — These Jobs Are First to Go

Jeannine Mancini

May 30, 2023, 12:42 pm

ARTIFICIAL INTELLIGENCE / TECH / AUGMENTED REALITY

Levi's will test AI-generated clothing models to 'increase diversity'

/ The denim tech will be u human mode completely re

th Lalaland's realistic AI-generated models later this & Co. / Lalaland.ai

By JESS WEATHERBED

Mar 27, 2023, 6:57 PM GMT+

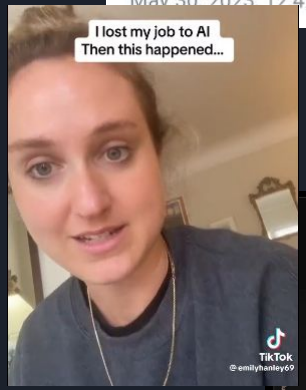
TECH • A.I.

The A.I. job culling has already begun and 4,000 people lost work last month to the technology, according to a new report

White-collar jobs are most at risk.

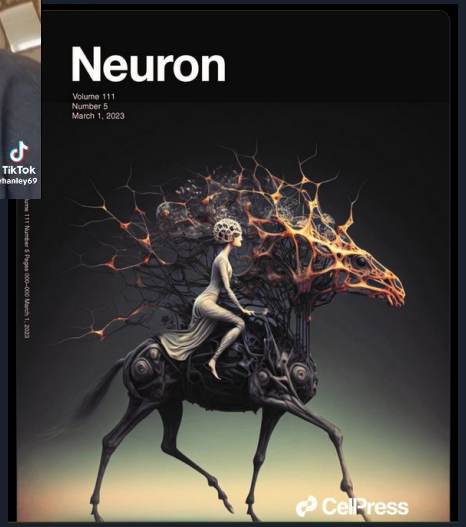
BY RACHEL SHIN

June 02, 2023 3:06 PM EDT



ChatGPT took their jobs. Now they walk dogs and fix air conditioners.

Technology used to automate dirty and repetitive jobs. Now, artificial intelligence chatbots are coming after high-paid ones.



Chinese AI News Anchor Works 24 Hours a Day, 365 Days a Year

MAR 17, 2023 PESALA BANDARA

EMERGENT TECH January 27, 2023

Buzzfeed to use ChatGPT for content creation after 12% employee layoff

VIDEO GAME NEWS

21ST MAR 2023 / 10:00 PM

Ubisoft is developing an AI tool 'that aims to support scriptwriters'

'UBISOFT GHOSTWRITER' WILL GENERATE NPC PHRASES AND SOUNDS

Posted by Tom Ivan

Other problems to consider

- Racism, biases and hallucinating
  - Data will always be more or less biased → Overwhelming amounts of data is of and by white people → Promoting CEO only shows pictures of white males in suits
  - Hallucinating - text generators feeding false information, making up lies on the spot
- Black box issue
  - When there is so much data that not even the engineers that build it knows what's going on - it's going too fast and it's too complex. They have no idea what's going on inside the machine.
- Environmental costs - ML leaves a huge carbon footprint
- "The Megafeed" - the oversaturated market where nothing is visible because of automation
  - AI's producing pictures, text and videos faster than humans - and uploading it
  - Get ready for not knowing if the things you see online are real or fake

Final note

- I want the rights to my FACE, my VOICE, my CONTENT!
- No one should be allowed to take my data, make money off of it, use it for scams or otherwise - especially not without explicit consent!
- I want AI companies to be held accountable for all the damage they are causing right now and in the future

I WANT DATA PROTECTION!