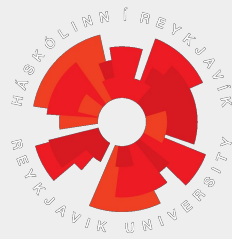


AI, Ethics, and Society

12.10.2026



About the Exam

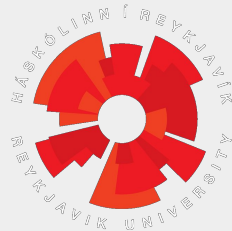
Regarding Reading Material

Everything in the reading material is assumed to be known.

However, I will not ask specific questions about reading material xyz.

But: Connecting answers to reading material helps you make your point.

If you give an answer that is related to some reading material in which an answer is provided and you argue the opposite, I will have to deduct points (assuming you do not take the reading material into account and show why you don't think it is correct).



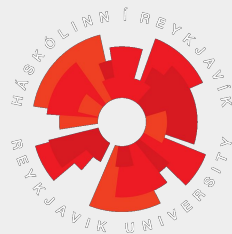
Recap

Last week we took a look at the EU AI Act and its Code of Practice

Four major parts:

1. Who has the obligation?
2. Who is the “User”?
3. Classification of AI according to its risks
4. General Purpose AI (GPAI)

We primarily focused on the classification and the Code of Practice for GPAI



Recap

Classification of Risks:

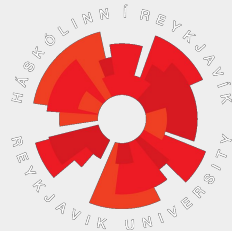
Four categories: 1) Unacceptable, 2) High-risk, 3) Limited risk, and 4) Minimal risk

Minimal: No regulations

Limited: Very few regulations (primarily show that AI is/was used)

High: All sorts of regulations (we took a closer look at that)

Unacceptable: Forbidden



Recap

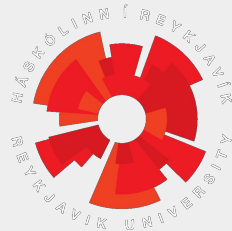
Code of Practice (CoP)

Used as guidelines for deploying GPAI systems.

Provides detailed measures that, if implemented, fulfill the obligations outlined in the EU AI Act for GPAI.

Not mandatory, you can fulfill the obligations differently.

But: You have to fulfill them. Signed CoP or not.



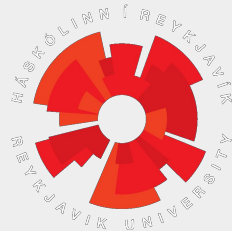
Questions

Definition of GPAI:

In Chapter 1 Article 3 of the EU AI Act you can find all definitions. Including number 63 on GPAI:

63. 'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market;

<https://artificialintelligenceact.eu/article/3/#p-article-3-63>

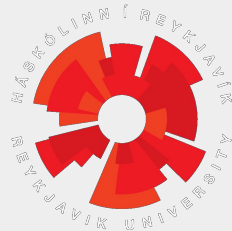


Questions

What is voluntary what is not?

The CoP (for GPAI) is voluntary, the EU AI Act is not. So all obligations outlined in the EU AI Act are legally binding. The way how you implement measures to fulfill them is not.

The CoP simply provides one way that, if implemented, gives a “guarantee” that all obligations of the EU AI Act regarding GPAI are fulfilled.

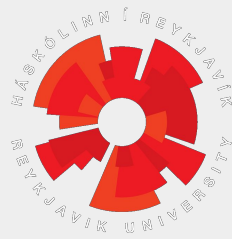


Questions

If the CoP is not mandatory, how can regulators ensure that all obligations are actually met?

The EU AI Office was founded for exactly that reason. They get access to evaluate if the measures introduced by the companies are sufficient.

How exactly that looks like? I really don't know.



Questions

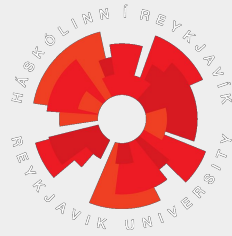
Is the EU AI Act any good? (Paraphrased a bunch of different questions into one)

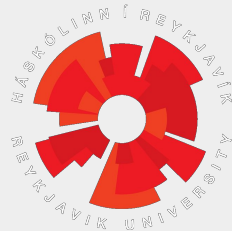
I honestly don't know. On the one hand, it doesn't go far enough, while on the other hand it is making it harder to build AI in the EU. So one could argue it is a loose-loose.

However, no regulation would be worse! At least some of the aspects are better than nothing. And while regulations always hinder innovation, I don't want unregulated innovation in the first place. If the regulation makes your innovation impossible, then maybe we don't want this innovation on the market.

And no: The market is not (necessarily) a good indicator for ethical practices. We will see that today.

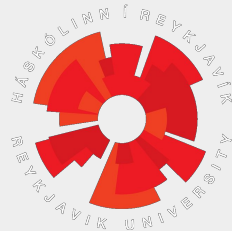
Open Questions?





AI in Practice

12.10.2026



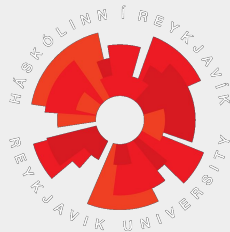
AI in Practice: Some Examples

Last year, we looked at some examples of what can go wrong and what can go right in AI development.

This year, I updated (some of) these slides to show how this changes (or does it change at all?)

I used news articles (not peer reviewed papers) for this. I linked every article in the slides, so if you want to know more, please take a look.

Also, please excuse the clickbait headings, that's apparently how news are today...



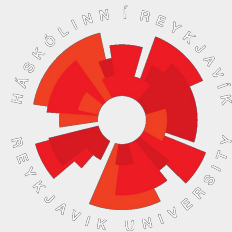
Examples of what can go wrong

Naftali Wambalo: I looked at people being slaughtered, people engaging in sexual activity with animals. People abusing children physically, sexually. People committing suicide.

Lesley Stahl: All day long?

Naftali Wambalo: Basically- yes, all day long. Eight hours a day, 40 hours a week.

<https://www.cbsnews.com/news/labelers-training-ai-say-theyre-overworked-underpaid-and-exploited-60-minutes-transcript/>



Examples of what can go wrong

Naftali Wambalo: I looked at people being slaughtered, people engaging in sexual activity with animals. People abusing children.

Fasica: I was told I was going to do a translation job.

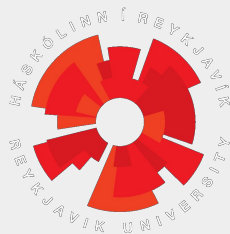
Lesley Stahl: All day long.

Lesley Stahl: Exactly what was the job you were doing?

Naftali Wambalo: Basically.

Fasica: I was basically reviewing content which are very graphic, very disturbing contents. I was watching dismembered bodies or drone attack victims. You name it. You know, whenever I talk about this, I still have flashbacks.

<https://www.cbsnews.com/news/labelers-training-ai-say-theyre-overworked-underpaid-and-exploited-60-minutes-transcript/>



Examples of what can go wrong

Naftali Wambalo: I looked at people being slaughtered, people engaging in sexual activity with animals. People abusing

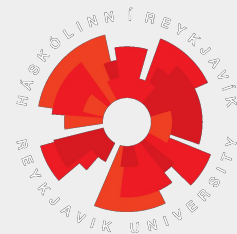
Fasica: I was told I was going to do a translation job.

To build that safety system, OpenAI took a leaf out of the playbook of social media companies like Facebook, who had already shown it was possible to build AIs that could detect toxic language like hate speech to help remove it from their platforms. The premise was simple: feed an AI with labeled examples of violence, hate speech, and sexual abuse, and that tool could learn to detect those forms of toxicity in the wild. That detector would be built into ChatGPT to check whether it was echoing the toxicity of its training data, and filter it out before it ever reached the user. It could also help scrub toxic text from the training datasets of future AI models.

you were doing?

ent which are very graphic, very disturbing contents. I was
e attack victims. You name it. You know, whenever I talk about

[erworked-underpaid-and-exploited-60-minutes-transcript/](#)



Examples of what can go wrong

Naftali Wambalo: I looked at people being slaughtered, people engaging in sexual activity with animals. People abused

Fasica: I was told I was going to do a translation job.

To build that safety system, OpenAI took a leaf out of the playbook of social media companies like Facebook, who had already shown it was possible to build

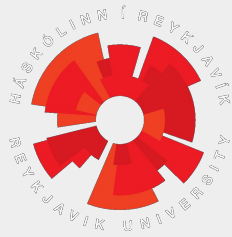
you were doing?

AI's that could detect toxic language like hate speech to help remove it from their platforms. The premise was simple: feed an AI with labeled examples of violence, hate speech, and sexual abuse, and that tool could learn to detect those forms of toxicity in the wild. That detector would be built into ChatGPT to check whether it was echoing the toxicity of its training data, and filter it out before it ever reached the user. It could also help scrub toxic text from training datasets of future AI models.

To get those labels, OpenAI sent tens of thousands of snippets of text to an outsourcing firm in Kenya, beginning in November 2021. Much of that text appeared to have been pulled from the darkest recesses of the internet. Some of it described situations in graphic detail like child sexual abuse, bestiality, murder, suicide, torture, self harm, and incest.

OpenAI's outsourcing partner in Kenya was Sama, a San Francisco-based firm that employs workers in Kenya, Uganda and India to label data for Silicon Valley clients like Google, Meta and Microsoft. Sama markets itself as an "ethical AI" company and claims to have helped lift more than 50,000 people out of poverty.

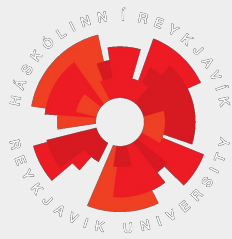
<https://time.com/6247678/openai-chatgpt-kenya-workers/>



But that was 2023, how about now?

Meta ended its contract with Sama after Kenyan data annotation workers told Swedish journalists they had viewed intimate footage, including people having sex, undressing, and using the toilet, captured by Meta's Ray-Ban smart glasses. The 1,108 workers received six days' notice. A class action lawsuit, UK and Kenyan regulatory investigations, and an EFF advisory followed. The case exposes the human infrastructure beneath AI: the workers who train the models see everything, own nothing, and lose their jobs when they talk about it.

<https://thenextweb.com/news/meta-smart-glasses-sama-kenya-workers>



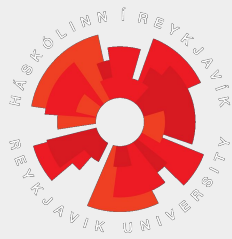
But that was 2023, how about now?

Meta ended its contract with Sama after Kenyan data annotation workers told Swedish journalists they had viewed intimate footage, including people having sex, undressing, and using the toilet, captured by Meta's Ray-Ban smart glasses.

The 1,108 workers received no compensation for their work, and regulatory investigations found that the company had not provided adequate human infrastructure benefits. The workers, who are mostly women, own nothing.

<https://thenextweb.com/r>

Sama was founded in 2008 as a social enterprise with the stated mission of providing dignified digital work to people in low-income communities. The company has operations in Kenya, Uganda, and India, and has provided data annotation services to some of the largest technology companies in the world, including Google, Microsoft, and Meta. The contract with Meta for smart glasses data annotation was one of several Sama held with the company. Workers were tasked with labelling images and video captured by the glasses to train Meta's AI models, a process that required them to view, categorise, and describe whatever the cameras had recorded.



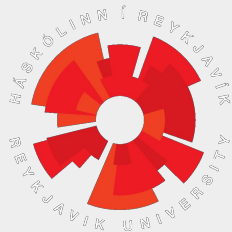
But that was 2023, how about now?

Meta ended its contract with Sama after Kenyan data annotation workers told Swedish journalists they had viewed intimate footage, including people having sex, undressing, and using the toilet, captured by Meta's Ray-Ban smart glasses.

The 1,108 workers received no compensation for their work. Sama was founded in 2008 as a social enterprise with the stated mission of providing dignified digital work to people in low-income communities. The company has operations in Kenya, Uganda, and India, and has provided data annotation

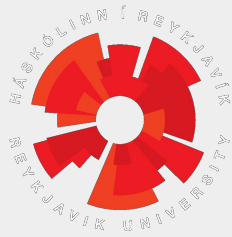
Kenyan data protection law. In the United States, the Clarkson Law Firm filed a class action lawsuit on behalf of consumers, alleging that Meta engaged in false advertising by marketing the glasses as "designed for privacy, controlled by you" while routing user footage through a human review pipeline in a country with weaker data protection enforcement than the markets where the glasses are sold. The Electronic Frontier Foundation published an advisory titled "Think

ies in the world, including for smart glasses data many. Workers were glasses to train Meta's AI e, and describe what-



Examples of what can go wrong

The IEA's special report *Energy and AI*, out today, offers the most comprehensive, data-driven global analysis to date on the growing connections between energy and AI. The report draws on new datasets and extensive consultation with policy makers, the tech sector, the energy industry and international experts. It projects that electricity demand from data centres worldwide is set to more than double by 2030 to around 945 terawatt-hours (TWh), slightly more than the entire electricity consumption of Japan today. AI will be the most significant driver of this increase, with electricity demand from AI-optimised data centres projected to more than quadruple by 2030.

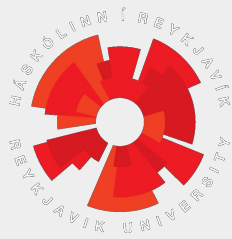


Examples of what can go wrong

The IEA's special report *Energy and AI*, out today, offers the most comprehensive, data-driven global analysis to date on the growing connections between energy and AI. The report draws on new datasets and extensive consultation with policy makers, the tech sector, the energy industry and international experts. It projects that electricity demand from data centres worldwide is set to more than double by 2030 to around 945 terawatt-hours (TWh), slightly more than the entire electricity consumption of Japan today. AI will be the most significant driver of growth in electricity demand from data centres, quadrupling by 2030.

Another energy security concern relates to the expanding demand for critical minerals used in the equipment in the data centres that power AI. The report provides first-of-its-kind estimates of demand from data centres for critical minerals, whose global supply is today highly concentrated.

<https://www.iea.org/news/ai-is-set-to-drive-surging-electricity-demand-from-data-centres-while-offering-the-potential-to-transform-how-the-energy-sector-works>



Examples of what can go wrong

The IEA's special report *Energy and AI*, out today, offers the most comprehensive, data-driven global analysis to date on the growing connections between energy and AI. The report draws on new datasets and extensive consultation with policy makers, the tech sector, the energy industry and international experts. It projects that electricity demand from data centres worldwide is set to more than double by 2030 to around 945 terawatt-hours (TWh), slightly more than the entire electricity consumption of Japan today. AI will be the most significant driver of demand from data centres.

qu Highlights:

- Data center developers are increasingly tapping into freshwater resources to quench the thirst of data centers, which is putting nearby communities at risk.
- Large data centers can consume up to 5 million gallons *per day*, equivalent to the water use of a town populated by 10,000 to 50,000 people.
- With larger and new AI-focused data centers, water consumption is increasing alongside energy usage and carbon emissions.
- Novel technologies like direct-to-chip cooling and immersion cooling can reduce water and energy usage by data centers.

Another energy security concern relates to the expanding demand for critical minerals used in the equipment in the data centres that power AI. The report

of demand from data centres for critical minerals, concentrated.

ht

[ring-the-potential-to-transform-how-the-energy-sector-works](#)

<https://www.eesi.org/articles/view/data-centers-and-water-consumption>

Examples of what can go wrong

The IEA's special report *Energy and AI*, out today, offers the most comprehensive, data-driven global analysis to date on the growing connections between energy and AI. The report draws on new datasets and extensive work in the tech sector, the energy industry and international organizations. It finds that electricity demand from data centres worldwide is set to rise to around 945 terawatt-hours (TWh), slightly more than the total electricity consumption of Japan today. AI will be the most significant driver of growth in electricity demand from AI-optimised data centres.

Highlights:

- Data center developers are increasingly tapping into local water resources, which is putting nearby communities at risk.
- Large data centers can consume up to 5 million gallons of water per day, a town populated by 10,000 to 50,000 people.
- With larger and new AI-focused data centers, water consumption and carbon emissions are rising.
- Novel technologies like direct-to-chip cooling and immersion cooling are being used by data centers.

Future planned capacity relative to current capacity. Western areas are **in bold**

	Percent projected growth	Future power capacity (MW)	Number of equivalent households that could be powered
Las Vegas / Reno	953	3,812	3,136,731
Salt Lake City	699	1,271	1,045,851
Phoenix	554	5,340	4,394,057
Atlanta	484	3,125	2,571,429
Dallas-Fort Worth	355	4,396	3,617,280
Chicago	239	2,773	2,281,783
Austin/San Antonio	237	1,993	1,639,954
Denver	196	311	255,909
Northern California	172	2,204	1,813,577
Northern Virginia	142	11,077	9,114,789
Houston	109	244	200,777
Los Angeles	50	398	327,497
Pacific Northwest	40	1,385	1,139,657
New York	39	264	217,234
New Jersey	37	781	642,651
Boston	0	160	131,657

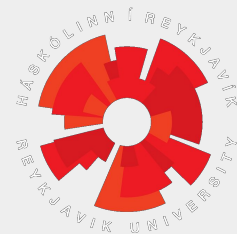
<https://www.eesi.org/articles/view/data-centers-and-the-environment>

<https://andthewest.stanford.edu/2025/thirsty-for-power-and-water-ai-crunching-data-centers-sprout-across-the-west/>

Examples of what can go wrong

Forensic Discipline	False Positive Rate	False Negative Rate
Facial recognition (FRT)	0.0001%	0.06%
Fingerprint comparison	0.1%	7.5%
Firearm comparison	1.01%	0.37%

Table 1: Example error rates for FRT ([27]) compared to fingerprint ([28]) and firearm ([29]) comparisons, as they are currently practiced. When applied to high-quality images, facial recognition technology (FRT) outperforms many traditional forensic disciplines in terms of accuracy. However, it is important to note that the reported error rates come from studies that have faced methodological criticism ([30]).



Examples of what can go wrong

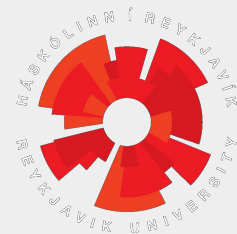
Forensic Discipline	False Positive Rate	False Negative Rate
Facial recognition (FRT)	0.0001%	0.06%
Fingerprint comparison	0.1%	7.5%
Firearm comparison		

Table 1: Example error rates for FRT ([27]) compared to fingerprint comparison, currently practiced. When applied to high-quality images, facial recognition outperforms traditional forensic disciplines in terms of accuracy. However, as shown in studies that have faced methodological criticism ([30]),

Our experiment finds that facial recognition technology (FRT) performance degrades under poor image conditions, particularly with blur, pose variation, and reduced resolution, and that this degradation is not evenly distributed across demographic groups. False positive and false negative rates increase with image degradation, disproportionately affecting individuals from marginalized race and gender groups. These disparities raise important concerns about the fairness and reliability of FRT when used in real-world law enforcement contexts, where image quality is often suboptimal. While this decision provides a conservative estimate of system performance under real-world conditions, it may further reduce accuracy compared to studies that exclude such images. As a result, our findings likely represent a lower bound on expected performance, particularly in operational settings where face detection is a prerequisite for matching.

Cuellar et al. (2025) “Accuracy and Fairness of Facial Recognition Technology in Low-Quality Police Images: An Experiment With Synthetic Faces”

Link: <https://arxiv.org/html/2505.14320v1>



Examples of what can go wrong

Forensic Discipline	False Positive Rate	False Negative Rate
Facial recognition (FRT)	0.0001%	0.06%
Fingerprint comparison	0.1%	7.5%
Firearm comparison		

Our experiment finds that facial recognition technology (FRT) performance degrades under poor image conditions, particularly with blur, pose variation, and reduced resolution, and that this degradation is not evenly distributed across demographic groups. False positive and false negative rates increase with image degradation, disproportionately affecting individuals from marginalized race and gender groups. These disparities raise important concerns about the fairness and

Table 1: Example error rates for FRT ([27]) compared to fingerprint comparison, currently practiced. When applied to high-quality images

LOCAL NEWS

Facial recognition technology error led to wrongful arrest, Brooklyn father says

CBS NEWS
NEW YORK

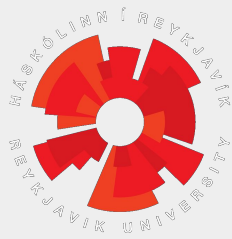
By Jennifer Bisram

Updated on: August 27, 2025 / 7:10 PM EDT / CBS New York



Police Images: An Experiment With Synthetic Faces”

<https://www.cbsnews.com/newyork/news/nyc-facial-recognition-technology-wrongful-arrest-indecent-exposure/>



Examples of what can go wrong

Forensic Discipline	False Positive Rate	False Negative Rate
Facial recognition (FRT)	0.0001%	0.06%
Fingerprint comparison	0.1%	7.5%
Firearm comparison		

Table 1: Example error rates for FRT ([27]) compared to fingerprint comparison, currently practiced. When applied to high-quality images, facial recognition technology performs better than fingerprint comparison. However, in real-world contexts, where image quality is often suboptimal. While this performance under real-world conditions, it may further reduce accuracy.

Our experiment finds that facial recognition technology (FRT) performance degrades under poor image conditions, particularly with blur, pose variation, and reduced resolution, and that this degradation is not evenly distributed across demographic groups. False positive and false negative rates increase with image degradation, disproportionately affecting individuals from marginalized race and gender groups. These disparities raise important concerns about the fairness and accuracy of FRT in law enforcement contexts, where image quality is often suboptimal. While this performance under real-world conditions, it may further reduce accuracy.

Facial recognition technology erred in wrongful arrest, Brooklyn father

CBS NEWS
NEW YORK

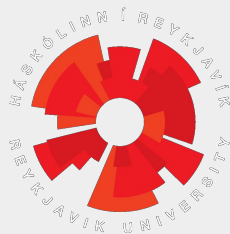
By Jennifer Bisram
Updated on: August 27, 2025 / 7:10 PM EDT / CBS New York

An attorney for Trevis Williams, 36, says back in April, the NYPD used facial recognition technology that misidentified Williams as a man who exposed himself to a woman in Manhattan earlier in the year. Williams spent two days in custody and was charged with an indecent exposure crime.

"He was arrested, he was prosecuted, and in spite of us telling the district attorney's office that we had evidence that Mr. Williams was nowhere near the scene of the crime at the time, the case continued," said Diane Akerman, an attorney with the Legal Aid Society's Digital Forensics Unit.

Williams' case was eventually dismissed.

<https://www.cbsnews.com/newyork/news/nyc-facial-recognition-technology-wrongful-arrest-indecent-exposure/>



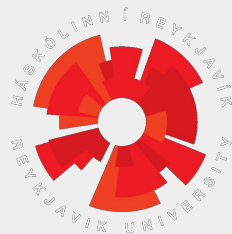
But that was 2025, how about now?

WHAT'S AT STAKE



On June 10, 2026, the ACLU and ACLU of Florida, with the law firm of Hoguet Newman Regal & Kenney, LLP, filed a wrongful arrest suit on behalf of Robert Dillon, a Florida man who was wrongfully arrested after police relied on an incorrect result from facial recognition technology.

<https://www.aclu.org/cases/dillon-v-city-of-jacksonville-beach>



But that was 2025, how about now?

WHAT'S AT STAKE



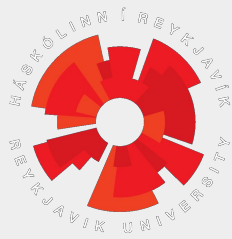
On June 10, 2026, the ACLU and ACLU of Florida LLP, filed a wrongful arrest suit on behalf of R... after police relied on an incorrect result from f...

<https://www.aclu.org/cases/dillon->

Facial recognition error prompts police to arrest Asian man for burglary 100 miles away

Exclusive: Alvi Choudhury claiming damages against Thames Valley police after biased technology confused him with man looking '10 years younger'

<https://www.theguardian.com/technology/2026/feb/25/facial-recognition-error-prompts-police-to-arrest-asian-man-for-burglary-100-miles-away>



But that was 2025, how about now?

WHAT'S AT STAKE



On June 10, 2026, the ACLU and ACLU of Florida, LLP, filed a wrongful arrest suit on behalf of Robert Williams, a Black man who was arrested after a facial recognition error identified him as a suspect in a 2019 murder case.

Facial recognition error prompts police to arrest Asian man for burglary

Lauren Yu,
She/Hers,
ACLU Speech, Privacy, and
Technology Project

Nathan Freed Wessler

Share This Page

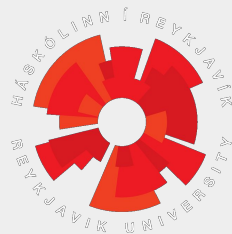


April 14, 2026

One ACLU client spent six months in jail, because police relied on facial recognition technology to incorrectly identify her as a suspect. She's the fourteenth person known to be wrongfully arrested due to the technology's failures.

When police arrested Kimberlee Williams, a grandmother living in Oklahoma, because of a warrant from Maryland, she was shocked. She had never been to Maryland in her life.

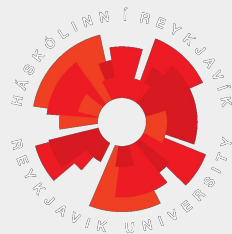
<https://www.aclu.org/news/privacy-technology/more-than-a-dozen-wrongful-arrests-due-to-police-reliance-on-facial-recognition-technology>



Examples of what can go wrong

**FCC PROPOSES \$6 MILLION FINE FOR ILLEGAL ROBOCALLS THAT
USED BIDEN DEEPPFAKE GENERATIVE AI VOICE MESSAGE**

*Steve Kramer Instigated Illegal Robocall and Spoofing Campaign Telling Voters Not to
Vote in 2024 New Hampshire Primary*



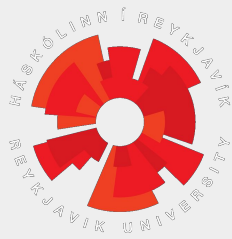
Examples of what can go wrong

FCC PROPOSES \$6 MILLION FINE FOR ILLEGAL ROBOCALLS THAT USED BIDEN DEEPPAKE GENERATIVE AI VOICE MESSAGE

Steve Kramer Instigated Illegal Robocalls

Two days before the New Hampshire 2024 presidential primary election, illegally spoofed and malicious robocalls carried a deepfake audio recording of President Biden's cloned voice telling prospective voters not to vote in the upcoming primary. Political consultant Steve Kramer was responsible for the calls and now faces a \$6 million proposed fine for perpetrating this illegal robocall campaign on January 21, 2024. The calls apparently violated the Truth in Caller ID Act by maliciously spoofing the number of a prominent local political consultant. The robocalls, made two days prior to the election, used a deepfake of President Biden's voice and encouraged voters to not vote in the primary but rather to "save your vote for the November election."

<https://docs.fcc.gov/public/attachments/DOC-402762A1.pdf>



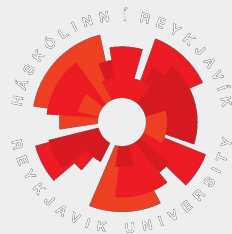
But that was 2024, how about now?

US MIDTERMS

AI deepfakes blur reality in 2026 US midterm campaigns

28/03/2026 – 11:10

<https://www.breakingnews.ie/world/ai-deepfakes-blur-reality-in-2026-us-midterm-campaigns-1880460.html>



But that was 2024, how about now?

US MIDTERMS

AI deepfakes blur reality in 2026 US midterm campaign

28/03/2026 – 11:10

<https://www.breakingnews.ie/world/ai-deepfakes-blur-reality-in-2026-us-midterm-campaigns-1880460>

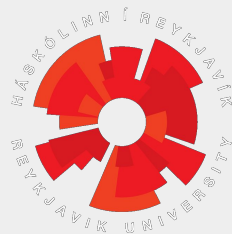
As the video opens, Democratic Texas State Representative **James Talarico** appears to stand in front of a Texas flag, beaming.

“Radicalised white men are the greatest domestic terrorist threat in our country,” the US Senate candidate seems to say into the camera.

As a voice whispers “white men,” Talarico continues: “So true. So true.”

But Talarico never filmed that video. Instead, the clip is an **AI-generated ad** from the National Republican Senatorial Committee (NRSC), the party’s Senate campaign arm, featuring a computer-altered Talarico reciting social media posts he wrote years ago.

The words “AI generated” show up in easy-to-miss font in the lower righthand corner.



But that was 2024, how about now?

US MIDTERMS

AI deepfakes blur real 2026 US midterm cam

28/03/2026 11:10

<https://kvia.com/politics/cnn-us-politics/2026/03/13/republicans-release-ai-deepfake-of-james-talarico-as-phony-videos-proliferate-in-midterm-races/>

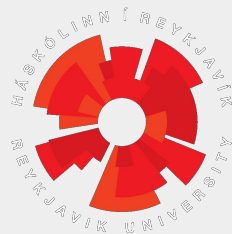
As the video opens, Democratic Texas State Representative **James Talarico** appears to stand in front of a Texas flag, beaming.

“Radicalised white men are the greatest domestic terrorist threat in our country,” the US Senate candidate seems to say into the camera.

As a voice whispers “white men,” Talarico continues: “So true. So true.”

A source familiar with the NRSC’s thinking described AI as a “consistently effective” way to highlight opposing candidates’ statements and said, “These are Talarico’s real words ... all we have done is visualize them for voters using a modern tool, within all legal and ethical parameters.” The source, however, said they did not have a comment to offer on the additional “Talarico” commentary the ad appears to have invented.

<https://kvia.com/politics/cnn-us-politics/2026/03/13/republicans-release-ai-deepfake-of-james-talarico-as-phony-videos-proliferate-in-midterm-races/>

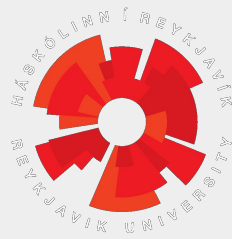


Examples of what can go wrong

When Zoom announced an update to its terms of service earlier this week that appeared to provide access to users' data for AI training, privacy advocates and customers rang the alarm.

"Zoom's [terms of service] now demand that they use AI to train on audio, face and facial movements, even private conversations without recourse, unconditionally and irrevocably," scientist Bryan Jones said in a tweet, "Opting out is not an option."

The backlash prompted Zoom to clarify its service terms in a blog post on Monday, in which it promised not to "use audio, video, or chat content for training our models without customer consent."



Examples of what can go wrong

When Zoom announced an update to its terms of service earlier this week that appeared to provide access to users' data for AI training, privacy advocates and customers rang the alarm.

"Zoom's [terms of service] now demand that they use AI to train on audio, face and facial movements even private conversations

Jones said in a tweet, "Opti

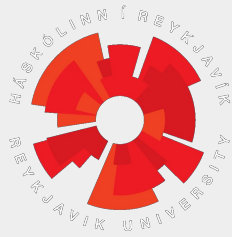
The backlash prompted Zc
promised not to "use audio

However, if a meeting host agrees to share data with Zoom, everybody participating in the meeting must share their data during that call.

This means participants who want their information to remain private must leave the Zoom call if their host consented to data-sharing. To be sure, this could be a problem for workers whose employers require them to attend Zoom sessions.

"If the administrator consents and it's your boss at your work who requires you to use Zoom, how is that really consent?" Katharine Trendacosta, director of policy and advocacy at the Electronic Frontier Foundation, told the Associated Press.

<https://www.cbsnews.com/news/zoom-privacy-issues-user-agreement/>

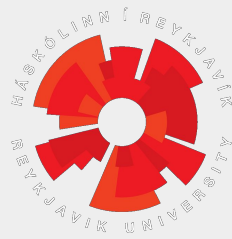


But that was then, how about now?

Meta is about to make your chats with its AI assistant part of its advertising machine, the company **announced** Wednesday. Beginning Dec. 16, conversations with Meta AI—the company’s chatbot embedded across Facebook, Instagram, Messenger, and even its new **Ray-Ban Display** smart glasses—will be used to determine which ads and recommendations show up in your feed.

The company will start formally notifying users of the change on Oct. 7. There’s no opt-out: If you don’t want your chatbot conversations influencing your ads, the only option is not to use Meta AI at all.

<https://fortune.com/2025/10/02/meta-ai-chatbot-update-exploits-privacy-monetize-chat-data-facebook-instagram-messenger-ray-ban-display-glasses/>



But that was then, how about now?

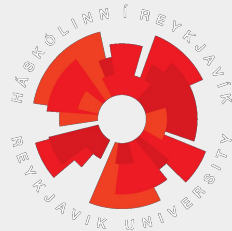
Meta is about to make your chats with its AI assistant part of its advertising machine, the company **announced** Wednesday. Beginning Dec. 16, conversations with Meta AI—the company’s chatbot embedded across Facebook, Instagram, Messenger, and even its new **Ray-Ban Display** smart glasses—will be used to determine which ads and recommendations show up in your feed.

The company will start formally notifying users of the change on Oct. 7. There’s no opt-out: If you don’t want your chatbot conversations influencing your ads, the only option is not to use Meta AI at all.

<https://fortune.com/2024/09/11/meta-ai-chatbot-ads-facebook-instagram/>

WASHINGTON, D.C. — A coalition of 36 privacy, consumer protection, children’s rights, and civil rights advocates today called on the Federal Trade Commission (FTC) to investigate and halt Meta’s recently announced plan to use conversations with its AI chatbots for advertising and content personalization. The [letter](#), sent to FTC Chair Andrew Ferguson and Commissioners, urges the agency to exercise its oversight authority and act under both Meta’s existing consent decree and Section 5 of the FTC Act to stop this practice from moving forward.

<https://epic.org/press-release-advocates-urge-ftc-to-halt-metas-plan-to-use-ai-chatbot-data-for-ads/>

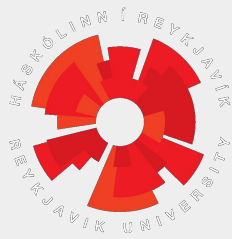


BUT

No rollout of these changes in the EU.

Data regulation laws do not allow this without consent or opt-out options (if at all).

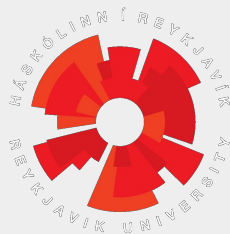
Regulations can affect these things!



BUT BUT

On May 7, 2026, the Council of the EU and the European Parliament announced their provisional agreement on targeted amendments to the EU AI Act, as part of the European Commission's Digital Omnibus initiative launched in late 2025, which proposed to delay the implementation of rules governing high-risk AI until 2027/28. The agreement acknowledges that AI Act compliance requires significant preparation and the introduction of supporting technical standards. The delay is intended to provide businesses with additional time to achieve compliance, while underscoring the expectation that implementation efforts should already be underway.

<https://www.hlc.com/en/publications/eu-legislators-agree-to-delay-for-highrisk-ai-rules>

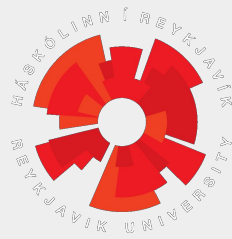


Examples of what can go wrong

New study raises concerns about AI chatbots fueling delusional thinking

First major study on 'AI psychosis' suggests chatbots can encourage delusions among vulnerable people

<https://www.theguardian.com/technology/2026/mar/14/ai-chatbots-psychosis>



Examples of what can go wrong

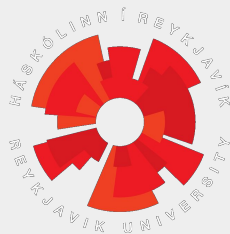
New study raises concerns about AI chatbots fueling delusional thinking

First major study on 'AI psychosis' suggests artificial intelligence may encourage delusions among vulnerable people.

<https://www.theguardian.com/r/14/ai-chatbots-psychosis>

A new scientific review raises concerns about how chatbots powered by artificial intelligence may encourage delusional thinking, especially in vulnerable people.

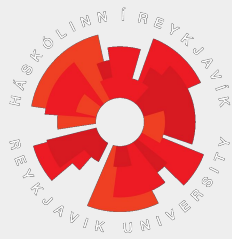
A summary of existing evidence on artificial intelligence-induced psychosis was published last week in **the Lancet Psychiatry**, highlighting how **chatbots** can encourage delusional thinking - though possibly only in people who are already vulnerable to psychotic symptoms. The authors advocate for clinical testing of **AI** chatbots in conjunction with trained mental health professionals.



BUT

In a statement, OpenAI said that ChatGPT should not replace professional mental healthcare, and that the company worked with more than 170 mental health experts to make GPT 5 safer. GPT 5 has still given problematic responses to prompts indicating mental health crises. OpenAI said it continues to improve its models with the help of experts.

<https://www.theguardian.com/technology/2026/mar/14/ai-chatbots-psychosis>

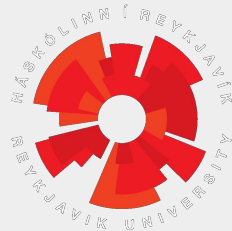


BUT

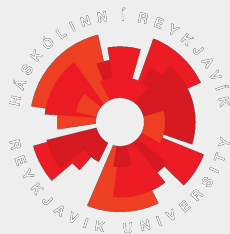
In a statement, OpenAI said that ChatGPT should not replace professional mental healthcare, and that the company worked with more than 170 mental health experts to make GPT-5 safer. GPT-5 has still given problematic responses to prompts continues to improve

<https://www.theguardian.com/technology/2024/03/14/openai-chatgpt-5-safer-but-still-gives-problematic-responses>

Creating effective safeguards for delusional thinking could be tricky, Morrin said, because “when you work with people with beliefs of delusional intensity, if you directly challenge someone and tell them immediately that they’re completely wrong, actually what’s most likely is they’ll withdraw from you and become more socially isolated.” Instead, it’s important to create a fine balance where you try to understand the source of the delusional belief without encouraging it - that could be more than a chatbot can master.



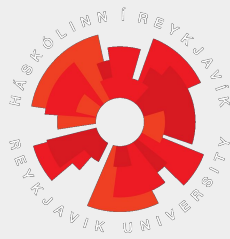
Let's also take a look at some benefits



Examples of what can go RIGHT

The NHS is launching the world's biggest trial of artificial intelligence to detect breast cancer, which could lead to faster diagnosis of the disease.

AI will be deployed to analyse two-thirds of at least 700,000 mammograms done in England over the next few years to see if it is as accurate and reliable at reading scans as a radiologist.



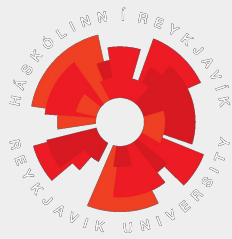
Examples of what can go RIGHT

The **NHS** is launching the world's biggest trial of artificial intelligence to detect breast cancer, which could lead to faster diagnosis of the disease.

AI will be deployed to analyse two-thirds of at least 700,000 mammograms done in England and Wales, with the aim of speeding up diagnosis and reducing the time at reading scans and waiting for results.

A **previous smaller study** of AI and breast cancer in Sweden in 2023, involving about 80,000 women, found that AI is safe to use, can reduce radiologists' workload by around half and does not lead to an increase in "false positives" - erroneous detections of symptoms of the disease.

<https://www.theguardian.com/society/2025/feb/04/nhs-to-launch-worlds-biggest-trial-of-ai-breast-cancer-diagnosis>



Examples of what can go RIGHT

The **NHS** is launching the world's biggest trial of artificial intelligence to detect breast cancer, which could lead to faster diagnosis of the disease.

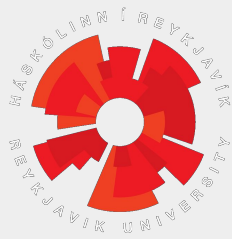
AI will be deployed to analyse two-thirds of at least 700,000 mammograms done in England at reading

The study, known as the **PRISM Trial** (Pragmatic Randomized Trial of Artificial Intelligence for Screening Mammography), is supported by a \$16 million award from the Patient-Centered Outcomes Research Institute (PCORI). The study will involve hundreds of thousands of mammograms interpreted at academic medical centers and breast imaging facilities in California, Florida, Massachusetts, Washington and Wisconsin. The AI support tool studied will be Transpara by ScreenPoint Medical with clinical workflow integration provided by the Aidoc aiOS platform.

<https://www>

[cer-diagnosis](#)

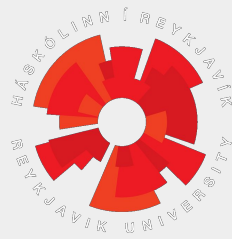
<https://www.uclahealth.org/news/release/ucla-lead-16-million-national-study-artificial-intelligence>



Update on this

There were fewer cases of breast cancer between two screening rounds, and of the cancers that did develop, fewer were advanced or aggressive. The final results from Lund University's MASAI trial are now available, and they show further benefits of AI-supported breast cancer screening. The study has already shown that AI support in mammography screening contributes to a 29 percent increase in detected breast cancers compared to traditional screening.

<https://www.lunduniversity.lu.se/article/ai-support-breast-cancer-screening-fewer-missed-cancer-cases>



Update on this

There were fewer cases of breast cancer between two screening rounds, and of the cancers that did develop, few were advanced or aggressive. The final results from the MASA trial are now available, and they show full support for AI-supported breast cancer screening. The study has shown that AI support in mammography screening led to a 29 percent increase in detected breast cancers compared to traditional screening.

"In the study, the research team has been able to demonstrate clinical safety, accuracy and effectiveness. Many regions in Sweden have already started to implement AI-supported mammography screening, and more are expected to join. It is a rather low bar to get started, in principle it is merely to integrate an AI software into existing IT systems," says Kristina Lång.

AI support has already begun to be implemented in other parts of Europe.

<https://www.lunduniversity.lu.se/article/ai-support-breast-cancer-screening-fewer-missed-cancer-cases>

Update on this

There were fewer cases of breast cancer between two screening rounds, and of the cancers that did develop, fewer were advanced or aggressive. The final results from the MASAI trial are now available, and they show that AI-supported breast cancer screening has been shown that AI support in mammography screening led to a 29 percent increase in cancer detection compared to traditional screening.

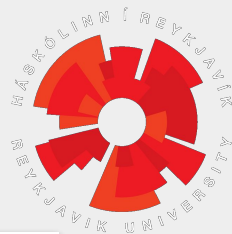
<https://www.lunduniversity.lu.se/press-releases/cancer-screening-fewer-missed-cancers>

standard double reading without AI. Use of AI resulted in a 12% reduction in interval cancer rate, meeting the criterion for non-inferiority. Use of AI also resulted in higher sensitivity (80.5% vs 73.8%) and the same specificity (98.5%) compared with standard double reading. Together with previous results from the MASAI trial,^{7,8} showing a 29% increase in cancer detection without an increase in false positives, and a 44% reduction in screen reading workload, these findings could motivate implementing AI in mammography screening.

"In the study, the research team has been able to demonstrate clinical safety, accuracy and effectiveness. Many regions in Sweden have already started to implement AI-supported mammography screening, and more are expected to join. It is a rather low bar to get started, in principle it is merely to integrate an AI system into the existing workflow," says Kristina Lång.

AI-supported screening has also been implemented in other parts of Europe.

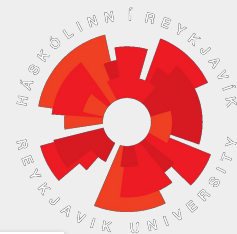
Gommers et al. (2026) - Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study: a randomised, controlled, non-inferiority, single-blinded, population-based, screening-accuracy trial (<https://www.sciencedirect.com/science/article/pii/S014067362502464X?via%3Dihub>)



Examples of what can go RIGHT

Generative AI has the potential to accelerate the adoption of **Universal Design for Learning (UDL) by offering multiple ways for students to access information, express ideas, and engage with course content**. This was reflected in findings from the **2025 NASPA Disability Knowledge Community meeting**, where disability services professionals described how AI is already being explored to support a wide range of academic needs. Many cited its value for improving note-taking and organizing information, assisting with task management, helping students complete assignments, and enhancing communication across settings.

<https://nationaldeafcenter.org/news-items/ai-as-an-accommodation-for-deaf-college-students/>



Examples of what can go RIGHT

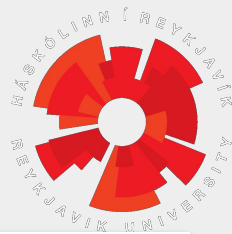
Generative AI has the potential to accelerate the adoption of **Universal Design for Learning (UDL)** by offering multiple ways for students to access information, express ideas, and engage with course content. This was reflected in findings from the **2025 NASPA Disability Knowledge Community meeting**, where disability services professionals described how AI is already being explored to support a wide range of student needs, from course management, helping students

<https://nationaldeafcenter.org/>

The Intersection of AI and Hearing Accessibility

In a crowded cafe, a cochlear implant user glances at their phone — not to check messages, but to read live captions of the conversation happening across the table. This everyday moment reflects a profound shift: artificial intelligence (AI) is reshaping hearing accessibility in ways that were once the realm of science fiction.

<https://www.innocaption.com/recentnews/the-intersection-of-ai-and-hearing-accessibility>



Examples of what can go right

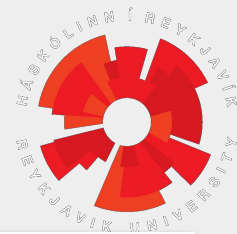
AI Turns Drone Footage Into Disaster Response Maps In Minutes

A system developed at Texas A&M uses drone imagery and artificial intelligence to rapidly assess damage after hurricanes and floods, offering life-saving insights in minutes.

Jul 28, 2025 | By Lesley Henton, Texas A&M University Division of Marketing and Communications

🕒 4 min. read

<https://stories.tamu.edu/news/2025/07/28/ai-turns-drone-footage-into-disaster-response-maps-in-minutes/>



Examples of what can go right

AI Turns Drone Footage Into Disaster Response Maps In Minutes

A system developed at Texas A&M uses drone imagery and artificial intelligence to rapidly assess damage after hurricanes and floods, offering life-saving insights in minutes.

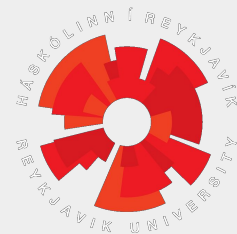
Jul 28, 2025 | By Lesley Henton,

<https://stories.tamu.edu>

15. Conclusions

In this article, we review the general state-of-the-art studies on wildfire management with AI. AI has high potential to enhance all disaster management, including floods, snowstorms, or earthquakes (Dimitrios [239]). The impact is not limited to predicting future events or creating high-resolution maps but can also help by preparing buildings and future constructions to limit damage (Parekh et al. [240] and Huang et al. [241]). Despite the large amount of research and proposals, recent events such as the Los Angeles Fire of 2024 tend to show that there is a lack of real-world application. This may be due to a lack of collaboration between researchers, developers, and field professionals.

<https://www.mdpi.com/2673-2688/6/10/253>

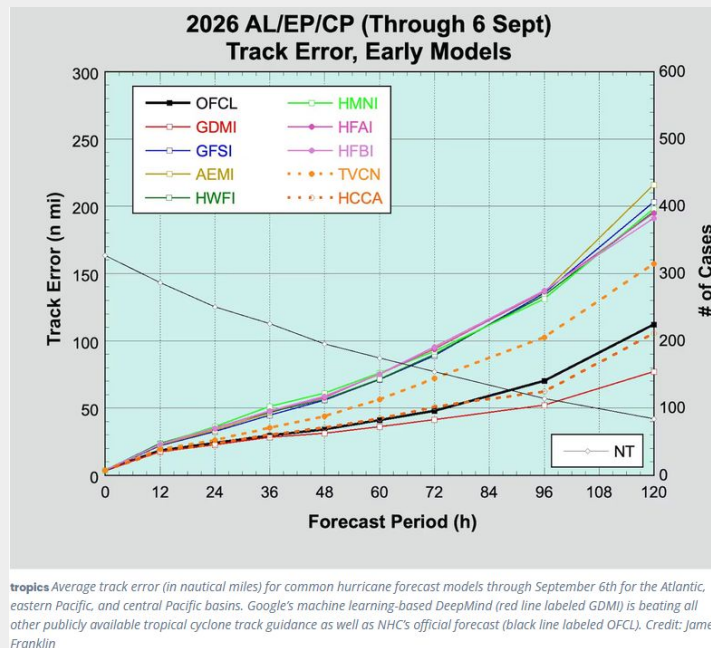


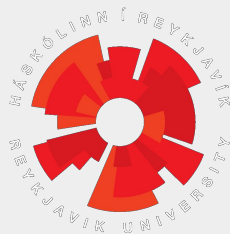
Examples of what can go right

The AI hurricane model that continues to blow away the competition

For the second straight hurricane season, Google's DeepMind is topping even official National Hurricane Center predictions

<https://www.local10.com/weather/hurricane/2026/09/10/the-ai-hurricane-model-that-continues-to-blow-away-the-competition/>





Examples of what can go right

While humans have assembled a lot of weather data, flash floods are too short-lived and localized to be measured comprehensively, the way the temperature or even river flows are monitored over time. That data gap means that deep learning models, which are increasingly capable of forecasting the weather, aren't able to predict flash floods.

To solve that problem, Google researchers used Gemini — Google's large language model — to sort through 5 million news articles from around the world, isolating reports of 2.6 million different floods, and turning those reports into [a geo-tagged time series](#) dubbed "Groundsource." It's the first time that the company has used language models for this kind of work, according to Gila Loike, a Google Research product manager. The research and dataset was [shared publicly](#) Thursday morning.

<https://techcrunch.com/2026/03/12/google-is-using-old-news-reports-and-ai-to-predict-flash-floods/>

With Groundsource as a real-world baseline, the researchers [trained a model](#) built on a Long Short-Term Memory (LSTM) neural network to ingest global weather forecasts and generate the probability of flash floods in a given area.

Google's flash flood forecasting model is now highlighting risks for urban areas in 150 countries on the company's [Flood Hub](#) platform, and sharing its data with emergency response agencies around the world. António José Beleza, an emergency response official at the Southern African Development Community who trialed the forecasting model with Google, said it helped his organization respond to floods more quickly.