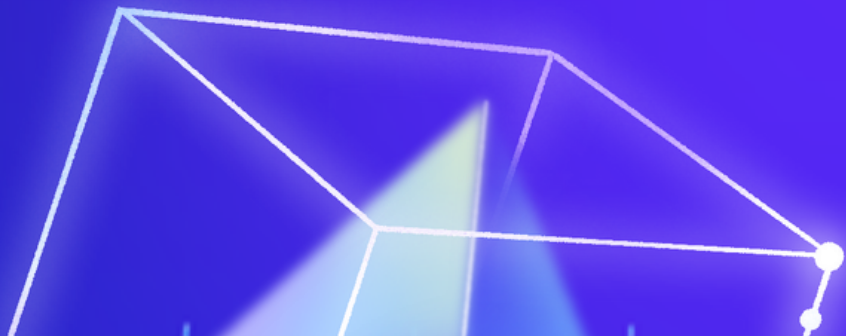


AI IS HERE, AND IT'S  
EVERYWHERE. HOW DO WE KNOW  
WHAT'S REAL AND WHAT'S  
NOT?

Alanna Burke  
DrupalCamp New Jersey  
March 15 2024



# Alanna Burke (she/her)



Community Manager, Developer  
Relations, and Documentation



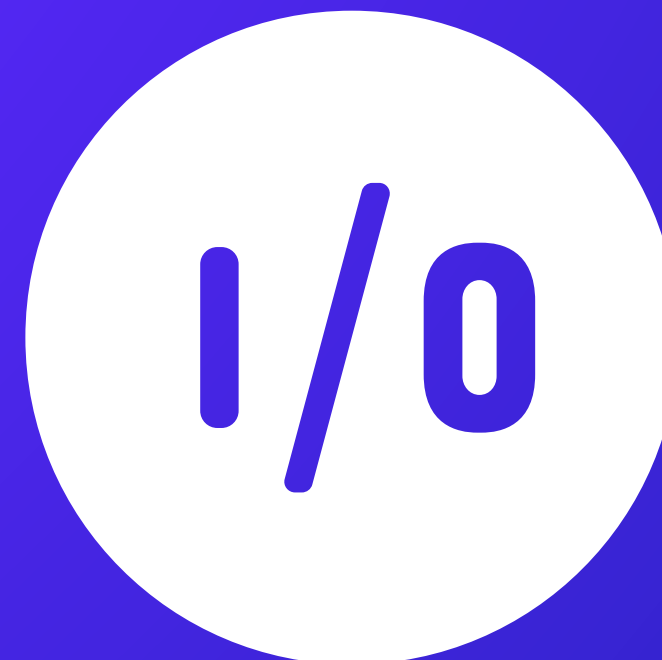
aburke626



alannaburke@fosstodon.org



alannaburke



**amazee.io**

part of Mirantis

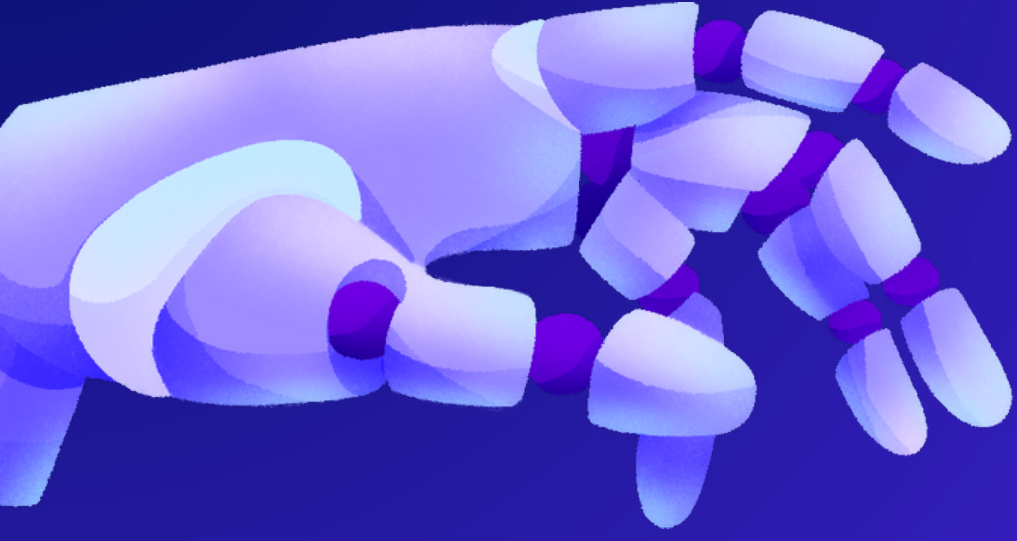






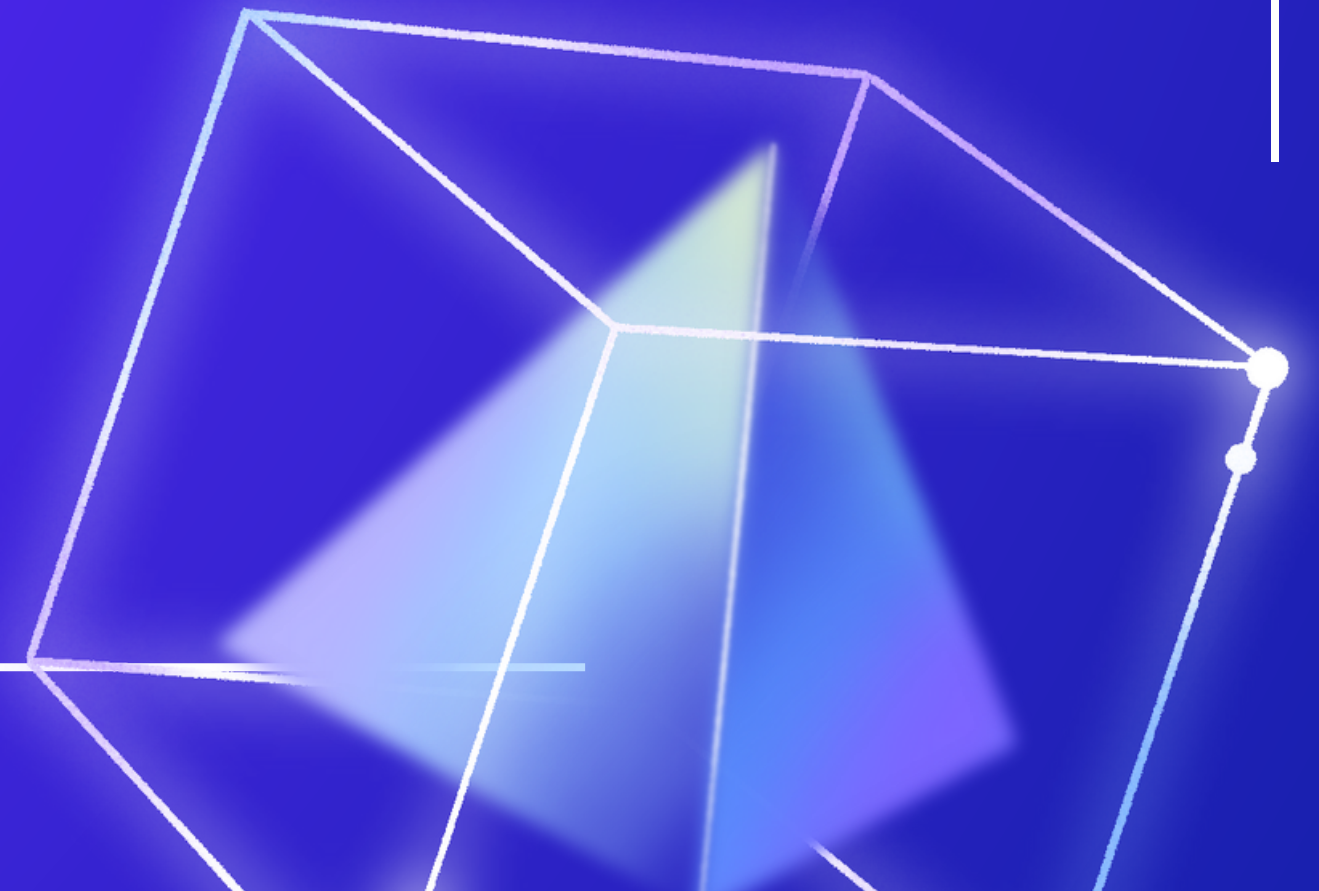






# OVERVIEW

- What is AI?
- Why is it important to distinguish between AI and human content?
- Deepfakes
- Telling the Difference
- Examples
- AI Content in the News
- Making AI Content Better





WHAT IS AI?



# Generative AI

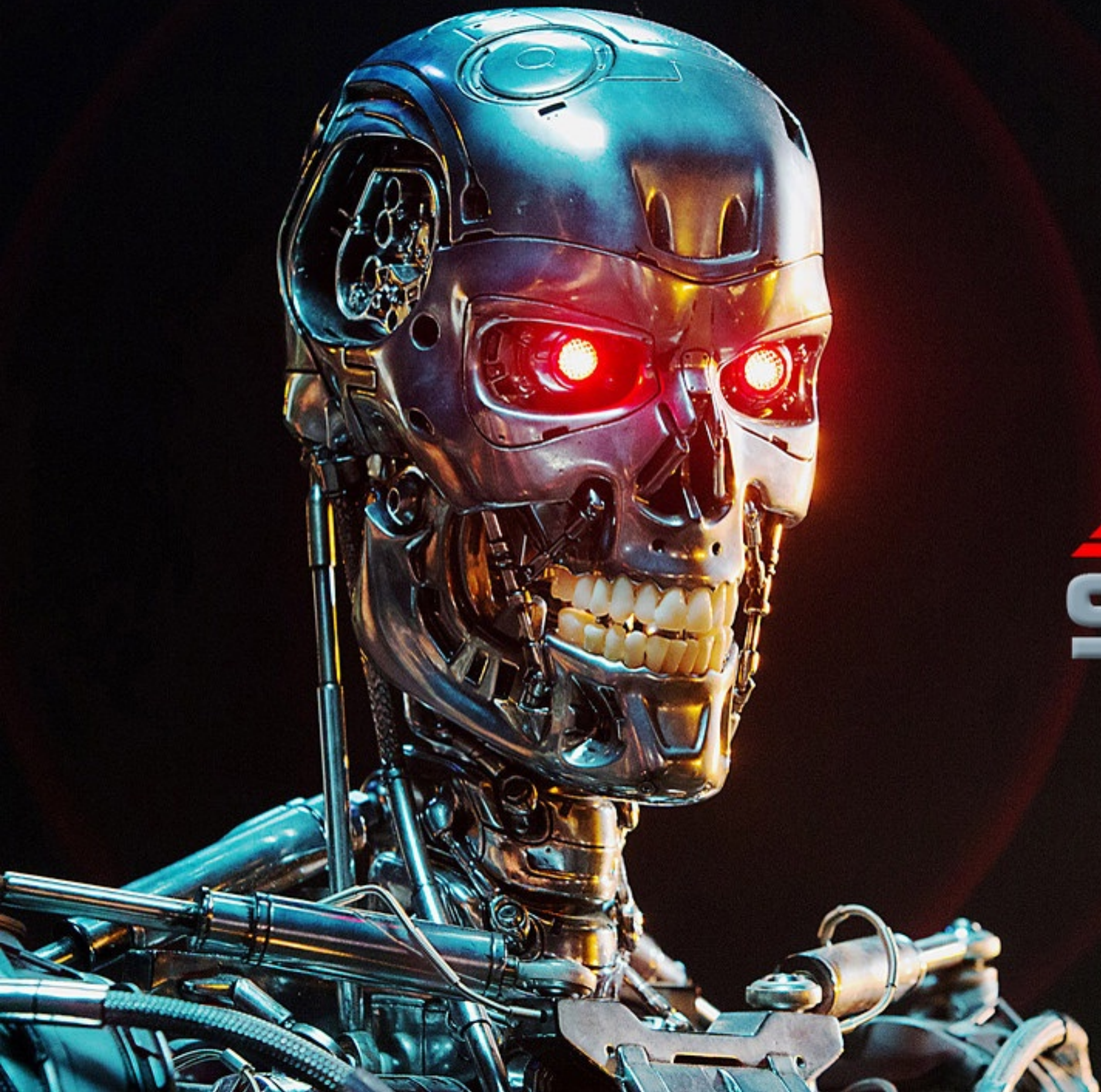
- Text, imagery, audio, and synthetic data
- Marketing content, books, videos, illustrations, voiceovers, articles, customer service chats, music, Willy Wonka Experiences
- This technology has been around since the 1960s but has advanced significantly with the introduction of generative adversarial networks (GANs) in 2014.
- GANs enable generative AI to create authentic images, videos, and audio.
- Recent advancements in transformers and language models have played a critical role in making generative AI more accessible.
- Transformers allow for larger models to be trained on vast amounts of text data, resulting in more in-depth answers.
- Generative AI works by starting with a prompt, which can be in the form of text, images, videos, or other inputs.
- Neural networks have transformed generative AI by learning the rules from existing data sets, rather than relying on explicitly crafted rules.

# Generative AI

- The benefits of generative AI: automating content generation, reducing effort in responding to emails, improving technical query responses, creating realistic representations, summarizing complex information, simplifying content creation
- The history of generative AI dates back to the 1960s with early implementations like the Eliza chatbot.

# Generative AI

- Some proponents of AI believe that generative AI is a crucial step towards general-purpose AI and possibly consciousness.
- However, there is no consensus on when AI will gain consciousness.
- Some predict it could happen by 2045, while others believe it could be much further off, with AI potentially reaching the intelligence of a dog by 2048.



# Popular AI Content Generators

- ChatGPT, Jasper, AI-Writer, Lex for text generation
- Dall-E 2, Midjourney, and Stable Diffusion for image generation
- Amper, Dadabots, and MuseNet for music generation
- CodeStarter, Codex, GitHub Copilot, and Tabnine for code generation
- Descript, Listnr, and Podcast.ai for voice synthesis

# Other Types of AI Content

- Music
- Social media videos
- Books

AI  
OR  
PHOTOSHOP?







WHY IS IT IMPORTANT TO  
DISTINGUISH BETWEEN AI AND  
HUMAN CONTENT?



# Why is it important?

- It's annoying
- Disinformation, scams
- Fake news
- Provenance
- Upholding ethical standards
- Protecting intellectual property rights

# DEEPFAKES





WHAT ARE DEEPFAKES?





WHY ARE DEEPFAKES  
DANGEROUS?





# DEEPPFAKE LAWS



# Deepfake Laws

- H.R.5586 – DEEPFAKES Accountability Act
  - “Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2023” or the “DEEPFAKES Accountability Act”.
  - To protect national security against the threats posed by deepfake technology and to provide legal recourse to victims of harmful deepfakes.





# Deepfake Laws

- **California:** In 2020 the state passed a law allowing victims of deepfake pornography to sue those who create and distribute sexually explicit deepfake material if the victim did not consent to it.
- **Florida:** In 2022 the state passed a law that prohibits the dissemination of sexually explicit deepfake images without the victim's consent.
- **Georgia:** A 2020 state law banned the online dissemination of falsely created pornographic images or videos.
- **Hawaii:** In 2021 the state outlawed the intentional creation, disclosure, or threat of disclosure of nonconsensual sexually explicit deepfake images or videos.
- **Illinois:** The law allows anyone falsely depicted in sexually explicit images or videos to sue the person who created the material. Victims can sue for damages and, to maintain their privacy, can use a pseudonym in court.

# Deepfake Laws

- **Minnesota:** A 2023 state law made it illegal to create sexually explicit deepfakes and to use deepfakes to influence an election.
- **New York:** In 2023 the state banned the distribution of pornographic images made using artificial intelligence without the consent of the subject.
- **Texas:** In 2023 the state made it a Class A misdemeanor to create sexually explicit, nonconsensual deepfake videos.
- **South Dakota:** A 2022 law made it a Class 1 misdemeanor to create deepfake pornography of an unwilling victim.
- **Virginia:** The state law was passed in 2019, as part of an existing law relating to revenge porn. The update adds "falsely created videographic or still image."

# Detecting Deepfakes

Pay attention to the face.

- Cheeks and forehead
- Does the skin appear too smooth or too wrinkly?
- Is the age of the skin similar to the age of the hair and eyes?
- Eyes and eyebrows – do shadows fall where you expect?
- Glasses – Is there any glare or reflection? Does it move when the person moves?
- Facial hair or lack thereof – does it look real? Moles?
- Blinking – do they blink enough? Too much?
- Lip movements – do they look like they're lip syncing?

# Tools to Detect Deepfakes

- FakeCatcher
- Operation Minerva
- Deepfake-O-Meter



# TELLING THE DIFFERENCE BETWEEN AI AND HUMAN CONTENT



“AI is a science, not an art—but catching it out is the  
opposite.”

- PC Mag



# Detecting AI Images - People

- Fingers - AI is getting better at this, but hands will still often have too many fingers.
- Hair - look especially at the edges of hair and where it intersects with other objects.
- Jewelry and accessories - check to see if jewelry and other accessories are accurately placed.
- Too-perfect skin - even compared to very filtered photos, AI just doesn't look right. The lighting and smoothness isn't realistic.
- Teeth - too many, too white, too straight.
- Mirrors - wrong reflection



# Detecting AI Images

- Vet it the way you would any other image
- Backgrounds
- “How the %^&\* did they do that?” if it seems totally impossible, it is
- Misaligned architecture, oddly curved walls, or unrealistic lighting.
- Sharp, overly rendered appearance
- Garbled text
- Illogical pieces of the image





Computers don't know how to make things imperfect.



# Detecting AI Text

- Doesn't flow naturally
- Lots of big words
- patterns, repetitions, or identical sentence structures
- lack of coherence and natural transitions
- may neglect grammar and punctuation rules
- very general information, no specifics



# Tools for Detecting AI Images

<https://www.aiornot.com/>

<https://copyleaks.com/ai-content-detector>

<https://www.scribbr.com/ai-detector/>

<https://isitai.com/ai-image-detector/>

<https://contentatscale.ai/ai-image-detector/>

Hive Moderation

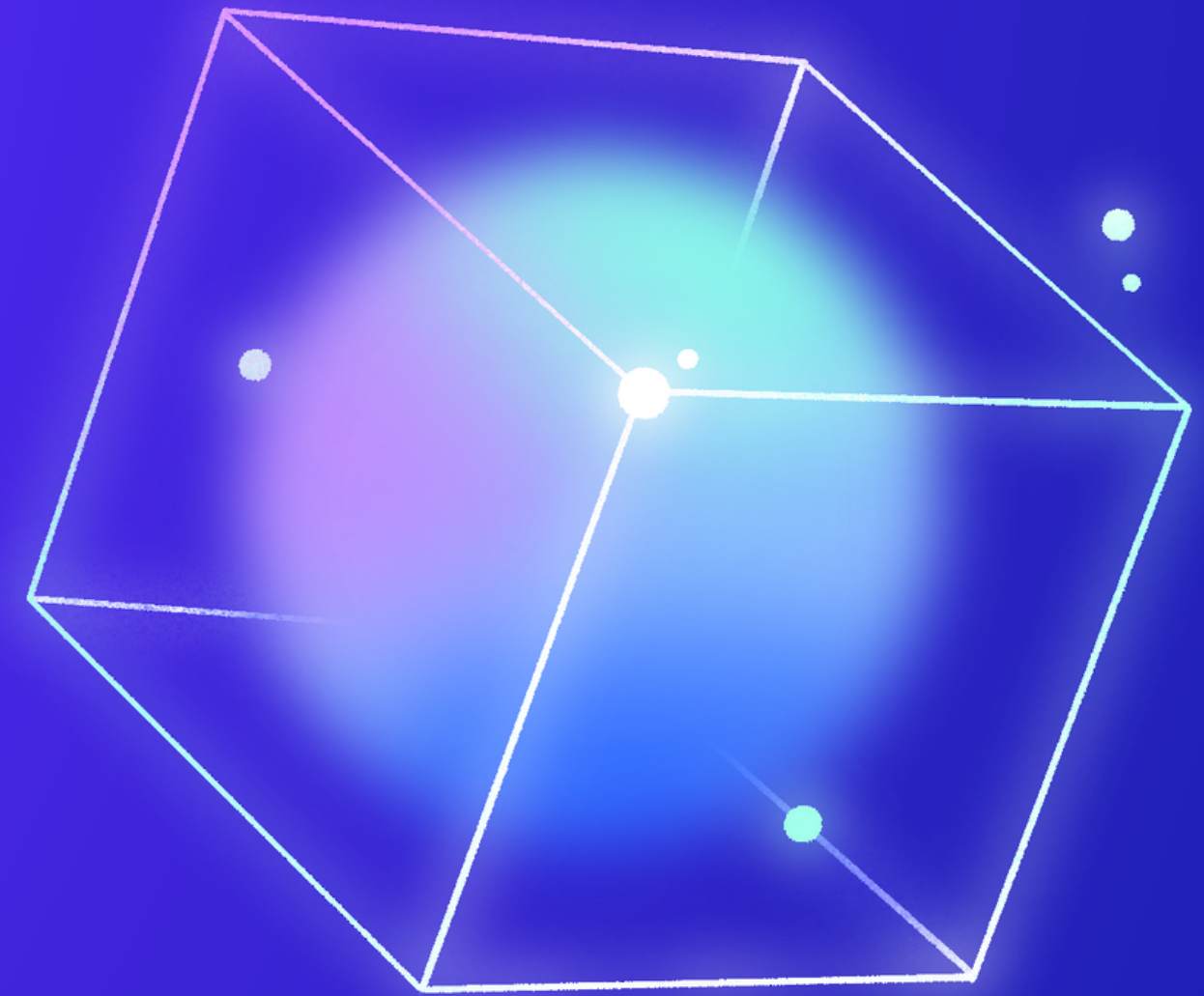
Hugging Face AI Detector

GPTZero

Originality AI



# EXAMPLES OF AI CONTENT



















































Travel Scotland Goals Group ·

Join

Henrik Brandt · 6d ·



Glasgow's annual pothole fishing contest took place on Bath street today as torrential rain pr... See more







Hello everyone, I am 122 years old, I made my own birthday cake with peach cream and filling, I started decorating cakes from 5 years old, I love it, and I can't wait to grow my baking journey 🥰❤️❤️❤�🍀



👍❤️ 123K

22K comments • 4.6K shares

👍 Like

💬 Comment

➦ Share





The background is a dark blue gradient. It features several abstract geometric elements: a white curved line with a dot at its end in the upper left; a white curved line with a dot at its end in the upper center; a white line with two dots in the upper right; a white wireframe cube in the lower left; and a large, curved, multi-colored shape in shades of purple and blue in the lower right.

# AI CONTENT IN THE NEWS

# AI Content in the News

- Willy Wonka Event
- Kate Middleton
- Emma Watson
- Marley's Mutts
- Taylor Swift selling cookware





MAKING AI  
CONTENT  
BETTER



# How can we make AI content better?

- AI Disclosure Act
  - To require generative artificial intelligence to disclose that their output has been generated by artificial intelligence, and for other purposes.
- White House Safety Accord



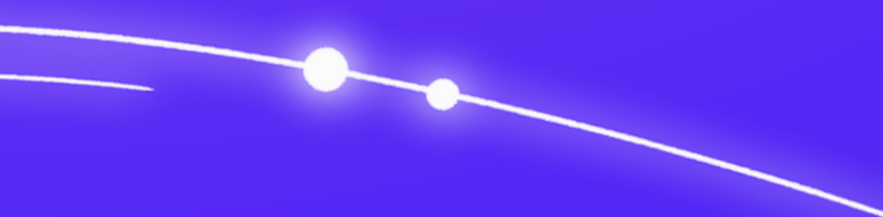
# Blueprint for an AI Bill of Rights

1. developing safety and security standards for AI
2. safeguarding privacy
3. promoting equity and civil rights
4. advocating for consumer, patient, and student protection
5. supporting workers
6. fostering innovation and competition
7. advancing US leadership in AI technology
8. ensuring responsible and effective government use of AI



# C2PA

The C2PA is a Joint Development Foundation project to collectively build an end-to-end open technical standard to provide publishers, creators, and consumers with opt-in, flexible ways to understand the authenticity and provenance of different types of media. C2PA opposes efforts to make content provenance measures mandatory.



# SynthID

- Developed by Google
- Watermarks audio tracks, undetectable by human ear even when audio is compressed, sped up or down, or has extra noise added.
- Watermark images by embedding directly into the pixels, undetectable by human eyes.
- Not foolproof against extreme manipulations.



# Meta

Facebook and its CEO, Nick Clegg, have announced a new approach to identifying and labeling AI-generated content on their platforms, including Facebook, Instagram, and Threads.



THANK YOU!



## References

- AI Content Detector*. (2024). Copyleaks. Retrieved March 15, 2024, from <https://copyleaks.com/ai-content-detector>
- AI Disclosure Act of 2023, H.R.3831, 118th Cong., 1st Sess. (2023)
- AI-Generated Image Detector*. (2024). Is It AI? Retrieved March 15, 2024, from <https://isitai.com/ai-image-detector/>
- AI or Not*. (2024). AI or Not. Retrieved March 15, 2024, from <https://www.aiornot.com/>
- Arctic\_Chilean. (2023, March 23). The 2001 Great Cascadia 9.1 Earthquake & Tsunami - Pacific Coast of US/Canada. [Online forum post]. Reddit. [https://www.reddit.com/r/midjourney/comments/11zyvlk/the\\_2001\\_great\\_cascadia\\_91\\_earthquake\\_tsunami/](https://www.reddit.com/r/midjourney/comments/11zyvlk/the_2001_great_cascadia_91_earthquake_tsunami/)
- Bartels, M. (2023, March 31). *How to Tell If a Photo Is an AI-Generated Fake*. Scientific American. Retrieved March 15, 2024, from <https://www.scientificamerican.com/article/how-to-tell-if-a-photo-is-an-ai-generated-fake/>
- Bartz, D., & Hu, K. (2023, July 21). *OpenAI, Google, others pledge to watermark AI content for safety, White House says*. Reuters. Retrieved March 15, 2024, from <https://www.reuters.com/technology/openai-google-others-pledge-watermark-ai-content-safety-white-house-2023-07-21/>
- Bond, S. (2023, June 13). *4 tips for spotting deepfakes and other AI-generated images : Life Kit*. NPR. Retrieved March 15, 2024, from <https://www.npr.org/2023/06/07/1180768459/how-to-identify-ai-generated-deepfake-images>
- Butuso, Ashtyn. [@lil\_buts]. (2024, March 10). I dk why but I brought the photo into photoshop and brightened it and the zipper is off too ok goodbye. [Image attached] [Post]. X. [https://twitter.com/lil\\_buts/status/1766956522267980089](https://twitter.com/lil_buts/status/1766956522267980089)
- Ciftci, U. A., Demir, I., & Yin, L. (2020, 7 15). FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 10.1109/TPAMI.2020.3009287



- Clegg, N. (2024, February 6). *Labeling AI-Generated Images on Facebook, Instagram and Threads* | Meta. Meta. Retrieved March 15, 2024, from <https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/>
- Coalition for Content Provenance and Authenticity. (2024). *C2PA: Overview*. C2PA. Retrieved March 15, 2024, from <https://c2pa.org/>
- Coilovercat. (2022, December 1). *How to identify scarily-accurate ai generated anime art from hand-drawn anime art (guide)*. [Online forum post]. Reddit. [https://www.reddit.com/r/AnimeSketch/comments/z9hmgf/how\\_to\\_identify\\_scarilyaccurate\\_ai\\_generated/](https://www.reddit.com/r/AnimeSketch/comments/z9hmgf/how_to_identify_scarilyaccurate_ai_generated/)
- Cole, S. (2024, February 23). *Romance Authors Riled by 'From Meet-Bot to AI-Do' Writing Workshop*. 404 Media. Retrieved March 15, 2024, from <https://www.404media.co/romance-authors-riled-by-from-meet-bot-to-ai-do-writing-workshop/>
- Collins, B. (2023, 10 14). *AI Or Not? How To Detect If An Image Is AI-Generated*. Forbes. Retrieved March 15, 2024, from <https://www.forbes.com/sites/barrycollins/2023/10/14/ai-or-not-how-to-detect-if-an-image-is-ai-generated/?sh=508eecbf3254>
- David, E., & Castro, A. (2023, September 12). *More companies commit to the White House AI safety accord*. The Verge. Retrieved March 15, 2024, from <https://www.theverge.com/2023/9/12/23870092/nvidia-ibm-adobe-white-house-ai-agreement-non-binding>
- David, E., & Castro, A. (2023, October 30). *Biden releases AI executive order directing agencies to develop safety guidelines*. The Verge. Retrieved March 15, 2024, from <https://www.theverge.com/2023/10/30/23914507/biden-ai-executive-order-regulation-standards>
- DeepFake Detection*. (n.d.). Papers With Code. Retrieved March 15, 2024, from <https://paperswithcode.com/task/deepfake-detection>
- DEEPFAKES Accountability Act, H.R.5586, 118th Cong., 1st Sess. (2023)

*Free AI Detector | Gemini, GPT4 and ChatGPT Detector.* (2024). Scribbr. Retrieved March 15, 2024, from <https://www.scribbr.com/ai-detector/>

Gillham, J. (2023, 8 25). *How To Detect AI-Generated Articles: A Step-By-Step Guide – Originality.AI.* Originality.ai. Retrieved March 15, 2024, from <https://originality.ai/blog/how-to-detect-ai-generated-articles>

Gowal, S., & Kohli, P. (2023, August 29). *Identifying AI-generated images with SynthID.* Google DeepMind. Retrieved March 15, 2024, from <https://deepmind.google/discover/blog/identifying-ai-generated-images-with-synthid/>

Hsu, T. (2023, June 28). *How Easy Is It to Fool A.I.-Detection Tools?* The New York Times. Retrieved March 15, 2024, from <https://www.nytimes.com/interactive/2023/06/28/technology/ai-detection-midjourney-stable-diffusion-dalle.html>

Hui, J. (2020, March 17). *Detect AI-generated Images & Deepfakes (Part 1).* Medium. Retrieved March 15, 2024, from <https://jonathan-hui.medium.com/detect-ai-generated-images-deepfakes-part-1-b518ed5075f4>

Hutson, M. (2023, January 6). *Detection Stays One Step Ahead of Deepfakes—for Now.* IEEE Spectrum. Retrieved March 15, 2024, from <https://spectrum.ieee.org/deepfake>

Intel. (2022, November 15). *Intel Introduces Real-Time Deepfake Detector.* Intel. Retrieved March 15, 2024, from <https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html#gs.r61ai2>

Intel. (2022, November 15). *Intel Introduces Real-Time Deepfake Detector.* Intel. Retrieved March 15, 2024, from <https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html#gs.r61ai2>

Irwin, K. (2024, February 7). *OpenAI Adds Labels to AI Images, But Says Metadata 'Easily' Removed.* PCMag. Retrieved March 15, 2024, from <https://www.pcmag.com/news/openai-labels-ai-images-metadata-easily-removed>

Jimenez, K., Weise, E., & Santucci, J. (2024, January 26). *Were Taylor Swift explicit AI photos illegal? Can she sue?* USA Today. Retrieved March 15, 2024, from <https://www.usatoday.com/story/news/nation/2024/01/26/was-deepfake-taylor-swift-pornography-illegal-can-she-sue/72359653007/>

Joint Development Foundation. (2024). Joint Development Foundation: Driving Adoption of Open Source and Standards. Retrieved March 15, 2024, from <https://jointdevelopment.org/>

Katzalcoatl. (2023, May 3). *A guide to detect scarily good AI art*. [Online forum post]. Tapas. <https://forums.tapas.io/t/a-guide-to-detect-scarily-good-ai-art/76024>

Klingler, N. (n.d.). *How to Detect AI-Generated Content - viso.ai*. Viso Suite. Retrieved March 15, 2024, from <https://viso.ai/deep-learning/ai-generated-content-detection/>

Knibbs, K. (2024, January 10). *Scammy AI-Generated Books Are Flooding Amazon*. WIRED. Retrieved March 15, 2024, from <https://www.wired.com/story/scammy-ai-generated-books-flooding-amazon/>

Koebler, J. (2024, January 8). *'Dogs Will Pass Away': Hackers Steal Dog Rescue's Facebook Page, Turn It Into AI Content Farm*. 404 Media. Retrieved March 15, 2024, from <https://www.404media.co/dogs-will-pass-away-hackers-steal-dog-rescues-facebook-page-turn-it-into-ai-content-farm/>

Koebler, J. (2024, January 24). *Iceland Has Its Own AI George Carlin Moment, Considers Law Against Deepfaking the Dead*. 404 Media. Retrieved March 15, 2024, from <https://www.404media.co/iceland-considers-laws-against-deepfaking-the-dead-after-ai-resurrects-beloved-comedian/>

Koebler, J. (2024, March 5). *Inside the World of TikTok Spammers and the AI Tools That Enable Them*. 404 Media. Retrieved March 15, 2024, from <https://www.404media.co/inside-the-world-of-tiktok-spammers-and-the-ai-tools-that-enable-them/>

Kolovos, B., & Taylor, J. (2024, January 30). *Georgie Purcell: Adobe casts doubt that Photoshop automation was to blame for Nine News editing of her image*. The Guardian. Retrieved March 15, 2024, from <https://www.theguardian.com/media/2024/jan/30/georgie-purcell-mp-photoshop-automated-image-nine-news-apology-victoria-animal-justice-party>

Lawton, G. (2024, 01 01). *What is Generative AI? Everything You Need to Know*. TechTarget. Retrieved March 15, 2024, from <https://www.techtarget.com/searchenterpriseai/definition/generative-AI>

Lu, Y. (2024, January 26). *No, That's Not Taylor Swift Peddling Le Creuset Cookware*. The New York Times. Retrieved March 15, 2024, from <https://www.nytimes.com/2024/01/09/technology/taylor-swift-le-creuset-ai-deepfake.html>

Mack, D. (2024, February 28). *Wonka Experience Oompa Loompa Kirsty Paterson Interview*. Vulture. Retrieved March 15, 2024, from <https://www.vulture.com/article/glasgow-sad-oompa-loompa-interview.html>

Maybe, M. (2022, October 14). *Can an AI learn to identify "AI art"?* Medium. Retrieved March 15, 2024, from <https://medium.com/@matthewmaybe/can-an-ai-learn-to-identify-ai-art-545d9d6af226>

McDonald, K. (2018, December 5). *How to recognize fake AI-generated images*. Medium. Retrieved March 15, 2024, from <https://kcimc.medium.com/how-to-recognize-fake-ai-generated-images-4d1f6f9a2842>

McLean, D. (2024, 2 12). *How to Detect AI Writing in 2024 (Expert Tips)*. Elegant Themes. Retrieved March 15, 2024, from <https://www.elegantthemes.com/blog/business/how-to-detect-ai-writing>

Naldrett, Ruby. [@rubynaldrett]. (2024, March 11). my analysis of the kate middleton photo saga is that they took her face from the vogue cover she did years ago and edited it in. [Image attached] [Post]. X. <https://twitter.com/rubynaldrett/status/1767184942859997400>

Novak, M. (2023, 3 27). *AI Creates Photo Evidence Of 2001 Earthquake That Never Happened*. Forbes. Retrieved March 15, 2024, from <https://www.forbes.com/sites/mattnovak/2023/03/27/ai-creates-photo-evidence-of-2001-earthquake-that-never-happened/?sh=42e51a153985>

OpenAI. (n.d.). *C2PA in DALL·E 3*. OpenAI Help Center. Retrieved March 15, 2024, from <https://help.openai.com/en/articles/8912793-c2pa-in-dall-e-3>

*Overview < Detect DeepFakes: How to counteract misinformation created by AI — MIT Media Lab*. (n.d.). MIT Media Lab. Retrieved March 15, 2024, from <https://www.media.mit.edu/projects/detect-fakes/overview/>

Porter, J. (2023, November 16). *Google is embedding inaudible watermarks right into its AI generated music*. The Verge. Retrieved March 15, 2024, from <https://www.theverge.com/2023/11/16/23963607/google-deepmind-synthid-audio-watermarks>

Prashnani, E., Goebel, M., & Manjuath, B. S. (2022, November 17). Generalizable Deepfake Detection with Phase-Based Motion Analysis. *ArXiv*, *abs/2211.09363*. <https://arxiv.org/abs/2211.09363>

Rohiin. (2024, March 10). Ai taking over Facebook. [Online forum post]. Reddit. [https://www.reddit.com/r/mildlyinfuriating/comments/1bbo0vq/ai\\_taking\\_over\\_facebook/](https://www.reddit.com/r/mildlyinfuriating/comments/1bbo0vq/ai_taking_over_facebook/)

Rowe, A. (2023, November 22). *9 Simple Ways to Detect AI Images (With Examples) in 2024*. Tech.co. Retrieved March 15, 2024, from <https://tech.co/news/ways-detect-ai-images-examples>

Savin, J. (2023, February 10). *Deepfake audio: Emma Watson's case is proof voice-cloning is a problem set to rise*. Cosmopolitan. Retrieved March 15, 2024, from <https://www.cosmopolitan.com/uk/reports/a42728762/emma-watson-deepfake/>

Souza, Pete [@petesouza]. (2024, March 14). I made this photograph of Prince George meeting President Obama in 2016. The digital file was “processed” with Photoshop, a [Photograph]. Instagram. <https://www.instagram.com/p/C4gCRWMLngj/>

Southern, M. G. (2023, August 29). *Google Unveils New Tool To Detect AI-Generated Images*. Search Engine Journal. Retrieved March 15, 2024, from <https://www.searchenginejournal.com/google-unveils-new-tool-to-detect-ai-generated-images/495082/>

Spyscape. (n.d.). *What are Deepfakes and How You Can Detect Them?*. Spyscape. Retrieved March 15, 2024, from <https://spyscape.com/article/what-are-deepfakes-and-how-you-can-detect-them-we-asked-an-ai-tool>

Steele, C. (2023, October 18). *The Internet Is Full of Deepfakes, and Most of Them Are Porn*. PCMag. Retrieved March 15, 2024, from <https://www.pcmag.com/news/the-internet-is-full-of-deepfakes-and-most-of-them-are-porn>

Steele, C. (2024, March 7). *How to Detect AI-Generated Images*. PCMag. Retrieved March 15, 2024, from <https://www.pcmag.com/articles/how-to-detect-ai-created-images>

Steele, C., & Mendes, B. (2024, January 29). *Speak Now: Only Taylor Swift Has the Power to Fight Deepfake Porn*. PCMag. Retrieved March 15, 2024, from <https://www.pcmag.com/opinions/speak-now-only-taylor-swift-has-the-power-to-fight-deepfake-porn>

Stouffer, C. (2023, November 1). *What are deepfakes? How they work + how to spot one*. Norton. Retrieved March 15, 2024, from <https://us.norton.com/blog/emerging-threats/what-are-deepfakes>

*This AI Image Detector Helps with Deepfake Detection*. (2024). Content @ Scale. Retrieved March 15, 2024, from <https://contentatscale.ai/ai-image-detector/>

Vernon, H. (2024, February 12). *What's Wrong With This Rental Listing? The Furniture Is AI*. VICE. Retrieved March 15, 2024, from

<https://www.vice.com/en/article/xgwpmk/ai-generated-furniture-real-estate-listings>

The White House OSTP. (2022, 10). *Blueprint for an AI Bill of Rights*. The White House. Retrieved March 15, 2024, from <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>