



PYMNTS®

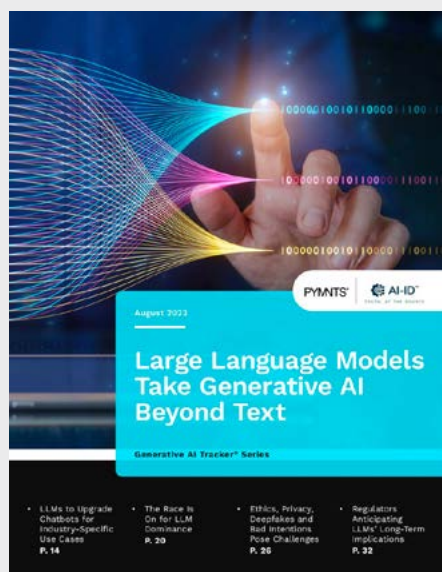


September 2023

Is That Content Generated by AI or Humans? Hard to Tell

Generative AI Tracker® Series

■ Read the previous edition



AUGUST 2023

**Generative AI
Tracker® Series**

- Companies Face the Dilemma of Labeling AI-Generated Text Appropriately **P. 10**
- Academia, Research and Publishing Adopt Content Scouts **P. 12**
- Tackling AI with AI May Not Be All It Is Cracked Up to Be **P. 16**
- Generative AI Renews Longtime Privacy Concerns **P. 24**

What's Inside

04 AI or Human? It Can Be Hard to Tell

Some studies find that more than half of individuals cannot accurately identify content made by AI chatbots such as ChatGPT.

10 Companies Face the Dilemma of Labeling AI-Generated Text Appropriately

Generative AI tools are promising but not foolproof. Their level of accuracy can be difficult to calculate and varies widely.

12 Academia, News, Publishers: Early Adopters of AI-Generated Content Detection

The effort to identify and label AI content is critical for industries that rely on authentic, trustworthy work to protect their integrity.

16 Tackling AI with AI May Not Be Ideal. Not Yet, at Least

Countermeasures to keep tabs on artificially generated content are AI tools themselves. They come with drawbacks.

24 Generative AI Renews Longtime Privacy Concerns

Artificially generated content, including deceptive and fraudulent content, is at our doorstep. Humans and companies cannot rein it in entirely.

30 About

Information on PYMNTS and AI-ID

PYMNTS®



Acknowledgment

The Generative AI Tracker® Series is produced in collaboration with AI-ID, and PYMNTS is grateful for the company's support and insight. [PYMNTS](#) retains full editorial control over the following findings, methodology and data analysis.

Background

AI or Human? It Can Be Hard to Tell

How do we decide whether a piece of text was created by a human, artificial intelligence (AI) or a combination of the two? Some studies find that more than half of individuals cannot accurately identify content made by [AI chatbots](#). That percentage jumps to 66% with content generated by GPT-4, the latest language model powering ChatGPT.

The implication for regulators, businesses and the public at large is that many are susceptible to [AI manipulation](#) without realizing it. This can have troubling and profound consequences for society, as AI content has the potential to influence large swaths of the population based on false premises.

Factors influencing ChatGPT's accuracy when generating responses:



Training data



Context



User input



Language complexity



Bias



Generative nature



Human evaluation

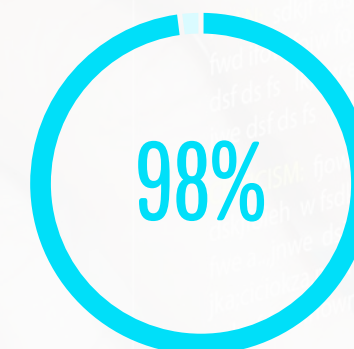
Background

The tools to spot AI-generated content are less than reliable.

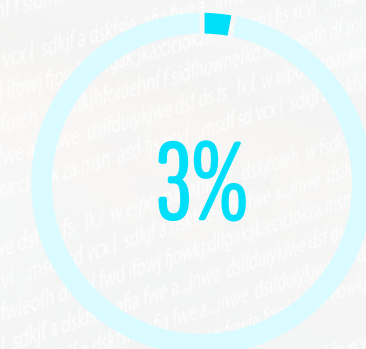
As generative AI grows, the countermeasures being developed to detect artificially generated content are close behind. Companies such as Originality.ai, Turnitin and Copyleaks are designing tools that discern between [unique content](#) and content compiled from online material.

These tools are promising but not infallible, as their accuracy can be difficult to measure and varies widely. In some cases, these solutions can be highly unreliable. OpenAI recently made news for retiring its AI classifier due to its “low rate of accuracy” in detecting artificial content.

ChatGPT’s level of accuracy can fluctuate significantly over time, in what experts call “drift.”



ChatGPT’s level of accuracy in March when identifying prime numbers



ChatGPT’s level of accuracy in June for the same task

Background

It is in the public interest to distinguish human- from AI-generated content.

As AI models grow in size and complexity, the ability to distinguish AI-produced content from human-sourced material becomes increasingly difficult. The same can be said for fraud: The more sophisticated the model, the more challenging it is to spot [fraudulent activity](#).

However, exposing artificial content is worth the effort and is critical to the integrity and survival of industries and organizations that depend on authentic, trustworthy information.



Companies of Note

Emerging Companies Are Tackling the Challenge of Spotting AI-Generated Content

A host of startups are cropping up to tackle detecting AI-generated content, and global organizations are fast on their heels. The emerging companies highlighted here represent just a few of the key players offering AI-generated content detection solutions. That list will only grow as the technology advances and new firms enter the market.

More and more universities, publishing companies, research organizations and financial firms are showing an interest in AI-generated content detection tools to ensure that the work of their students, writers, researchers and analysts is authentic, original and reliable.

GPTZero secured seed funding in the spring to offer [AI-generated content detection tools](#) using existing [generative AI models](#). Enterprises and educators can integrate these tools into their processes using application programming interfaces (APIs). Similarly, Copyleaks obtained seed funding in 2022 for its AI-generated content detection tool, a Chrome extension that identifies [synthetically generated content](#) and plagiarism.

Originality.ai is another content detection tool that seeks to recognize plagiarism and paraphrasing. Turnitin offers a similar solution for academic settings with the promise of delivering academic integrity by spotting plagiarism.

AI-generated content detection tools are far from failsafe. They will require iterative efforts to improve accuracy.



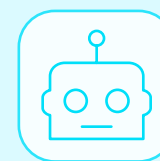
74%

Proportion of AI-generated text that OpenAI said its classifier tool missed detecting



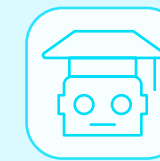
68%

Share of employees who do not disclose their use of generative AI to their employers



#10

How AI-generated content detection ranked in educational challenges in 2022



#1

How AI-generated content detection ranked in educational challenges in 2023

Innovation and Use Cases

Academia, News, Publishers: Early Adopters of AI-Generated Content Detection

Just as financial organizations prioritize fraud detection, industries and institutions that depend on authentic, trustworthy content to preserve their integrity are seeking technologies that can correctly and comprehensively identify AI-generated content. These sectors include academia, research and news organizations and publishers.

Schools and teachers want to make sure their students are doing their own work by determining who — or what — produced it. Research organizations value the integrity of original content submitted by their researchers and graduate students. News organizations pride themselves on producing accurate and original content. Such use cases are examples of how AI-generated content detectors can be applied to combat fraud and distinguish authenticity from counterfeits.



Innovation and Use Cases

AI can be used to produce inauthentic content or commit fraud. It can also be used to stop it.

AI is not inherently good or bad. It can, however, assume a good or bad face or serve helpful or harmful causes depending on who is wielding its power and for what purpose. Guiding principles and regulations such as the European Union's General Data Protection Regulation (GDPR) can provide valuable context to assess where they fall.

Just as generative AI can be used to replace human creativity in the industries where the latter is most essential, [fraudsters can use generative AI](#) to tap personal data, access financial funds, create fake identities or conduct fraudulent transactions. Generative AI models [lack transparency and governance guidelines](#), giving fraudsters the very shadows in which they seek to hide.

However, businesses, academics, fraud examiners and researchers can also leverage AI to boost their creativity, improve customer experience, automate processes, augment transaction monitoring and [spot artificial content, fraudulent transactions](#) and fraudsters. Financial institutions such as Visa, Bank of America and PayPal use generative AI as part of their efforts to stop illicit transactions and malicious actors.



IRENE SOLAIMAN
Policy director

“The bigger and more powerful the model, the harder it is to build AI models to detect what text is written by a human and what isn’t.”



Hugging Face

Source: Joe Marshall: As AI cheating booms, so does the industry detecting it: ‘We couldn’t keep up with demand’. The Guardian. July 5, 2023. <https://www.theguardian.com/technology/2023/jul/05/as-ai-cheating-booms-so-does-the-industry-detecting-it-we-couldnt-keep-up-with-demand>. Accessed August 2023

Issues and Challenges

Tackling AI with AI May Not Be Ideal. Not Yet, at Least

As generative AI grows by leaps and bounds, countermeasures to keep tabs on artificially generated content are emerging. The challenge for companies developing these solutions is not only that their tools are new, unproven and offer mixed success but also that these tools' capabilities will be increasingly stretched as AI-generated content grows and becomes more humanlike.

Leading AI developers and technology companies — including Amazon, Google, Meta, Microsoft and OpenAI — have made voluntary commitments to the Biden-Harris Administration to ensure that their generative AI content tracker tools are effective. Whether those businesses honor their commitments, and to what degree their promises hold substance, remains to be seen.

“You can’t trust anything
that you see or hear
[regarding the U.S.
elections, thanks to AI].”



ERIC SCHMIDT
Former CEO



Source: Bhaskar Chakravorti: The AI Regulation Paradox. Foreign Policy. August 4, 2023. <https://foreignpolicy.com/2023/08/04/ai-regulation-artificial-intelligence-democracy-elections/>. Accessed August 2023.

Issues and Challenges

Content scouts are AI tools themselves.

Generative AI tools use keywords related to a topic as input and scour the web for content with similar keywords. They then rewrite that content autonomously. This is where content scouts come in. These tools employ AI to analyze the text to detect certain patterns and characteristics indicative of AI-created content. No foolproof methods to detect and expose artificial content exist today. The current crop of content scouts — AI tools themselves — have a spotty track record in terms of accuracy and consistency. The reality is that there may never be a perfect tool, especially as AI-generated content becomes more sophisticated and humanlike.

Nevertheless, content scouts work to identify generative AI content based on [clues in the text](#), such as a lack of typos or misspellings, information with no accompanying sources or shorter, flat-sounding sentences — all dead giveaways. Repeated words or phrases and the absence of analysis or commentary offer additional hints.



SAMUEL LEVINE
Director



FEDERAL TRADE COMMISSION
PROTECTING AMERICA'S CONSUMERS

“The FTC welcomes innovation, but being innovative is not a license to be reckless. We are prepared to use all our tools, including enforcement, to challenge harmful practices in this area.”

Source: Cat Zakrzewski: FTC investigates OpenAI over data leak and ChatGPT's inaccuracies. Washington Post. July 13, 2023. <https://www.washingtonpost.com/technology/2023/07/13/ftc-openai-chatgpt-sam-altman-lina-khan/>. Accessed August 2023.

Issues and Challenges

Will voluntary commitments be enough?

If history is any indication, the initial measures technology companies and regulators have taken may not be enough. The preventive measures that technology companies offer today, either with good intentions or to generate goodwill, may also be inadequate. This is partly because of commercial interests and partly because of unforeseen societal implications of generative AI.

Going forward, companies that use generative AI will draw scrutiny from regulators. Regulators are likely to cite concerns around consumer or public manipulation, the spread of misinformation or dissemination of harmful content. In response, companies leveraging generative AI will most likely emphasize the benefits of these technologies, downplay troubling scenarios as outliers and seek ways to circumvent regulation.



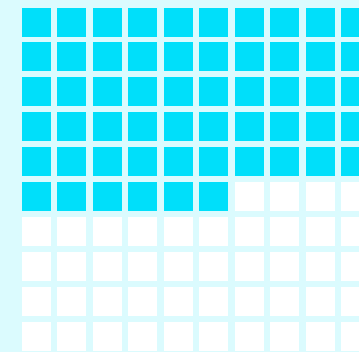
Issues and Challenges

Cybersecurity dons a layer of AI.

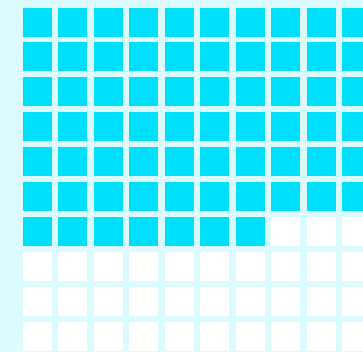
Cybersecurity breaches are a primary concern across many industries, prompting organizations to address the threat by enhancing their data security protocols to block unauthorized access and stop account takeovers. To this end, companies are using AI to integrate fraud monitoring, identity verification databases, data encryption, virtual private networks (VPNs) and antivirus software into their business systems and processes.

While the use of generative AI is currently in its early stages, it will not be long before AI companies incorporate generative AI capabilities into their solutions.

The repercussions of cybersecurity breaches for these industries include theft of trade secrets, leaks in confidential information and the proliferation of synthetic identities that can lead to the creation of fake accounts or takeover of valid, existing accounts. They can also hurt the company's reputation and brand in the public eye.



56%
**Share of business leaders
incorporating AI into their
cybersecurity protocols**



67%
**Share of business leaders
planning to invest in AI
technologies to boost
their cybersecurity
protocols**

Policy and Regulation

AI Detection Tools Can Be a Blessing and a Curse

Governing and regulating AI across borders and businesses have long-term and far-reaching implications — some unintended, some yet unknown. These regulations must not only address fraud, data privacy and security breaches but also balance the protection of free speech while simultaneously preventing misinformation.

In the near future, regulators will likely focus on more tangible concerns, such as consumer data protection, intellectual property law violations and financial fraud. In the long term, they will need to expand their efforts to deal with broader matters, including setting guardrails to protect global democratic institutions.



Policy and Regulation

Regulatory and business disclosures are coming soon.

Until AI detection tools are perfected and future-proofed, the public will have to rely on self-disclosures, voluntary initiatives and new legislation to ensure the accuracy, credibility and authenticity of AI systems. [Watermarking AI content](#) is another tool that can help. Government regulation, of course, will go a long way toward safeguarding consumers and the public, but it remains a work in progress.

Regulatory efforts such as the [REAL Political Advertisements Act](#), introduced in the U.S. House of Representatives in May, offers an example of how lawmakers are trying to rein in AI. If passed, the law will require advertisers to disclose any AI-generated content used in political ads. This is particularly relevant ahead of the 2024 presidential elections.



“Given the many risks of using personal data collected by AI tools, employers should vet the tools they use and negotiate service agreements to reduce their risk. In these service agreements, employers should consider requiring assurances that the generative AI service complied with applicable data protection laws when collecting and processing personal data.”

ZOE ARGENTO

Co-chair, privacy and data security practice

Littler

Source: Zoe Argento: Data protection issues for employers to consider when using generative AI. IAPP. August 9, 2023. <https://iapp.org/news/a/data-protection-issues-for-employers-to-consider-when-using-generative-ai/>. Accessed August 2023.

Policy and Regulation



Several AI laws and investigations, both in the U.S. and abroad, are in the works. The Federal Trade Commission is [investigating OpenAI](#) to determine whether ChatGPT violates customer protection laws or misuses personal data, as well as generates misleading or harmful information. At issue are OpenAI's data models.

How these investigations amount to regulation, fines and consequences remains a question mark. How and which entity will have the authority to oversee whether the models are retrained properly also remains to be seen.



\$8.8B

Total fraud losses in U.S. in 2022



€1.65B

Fines issued by European regulators for GDPR and **data violations related to AI**

About

PYMNTS® [PYMNTS](#) is where the best minds and the best content meet on the web to learn about “What’s Next” in payments and commerce. Our interactive platform is reinventing the way in which companies in payments share relevant information about the initiatives that shape the future of this dynamic sector and make news. Our data and analytics team includes economists, data scientists and industry analysts who work with companies to measure and quantify the innovation that is at the cutting edge of this new world.



AI-ID™ is a leading-edge provider of AI and Generative AI Platform Management, Regulation, Governance and Output control capabilities and solutions. With purpose-built solutions including the AIGX File, AI-ID™ Platform and AI Registry™ components, the patented AI-ID technology suite and services give Generative AI platforms, providers, their customers and all regulators and stakeholders the trust and data provenance needed in a rapidly advancing AI world. We are a company with a primary mission: solving the key challenges facing the AI marketplace today — while providing strong integrated foundational capabilities for its long-term future success. Learn more at www.ai-id.io.

We are interested in your feedback on this report. If you have questions or comments, or if you would like to subscribe to this report, please email us at feedback@pymnts.com.

Disclaimer

The Generative AI Tracker® Series may be updated periodically. While reasonable efforts are made to keep the content accurate and up to date, PYMNTS MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, REGARDING THE CORRECTNESS, ACCURACY, COMPLETENESS, ADEQUACY, OR RELIABILITY OF OR THE USE OF OR RESULTS THAT MAY BE GENERATED FROM THE USE OF THE INFORMATION OR THAT THE CONTENT WILL SATISFY YOUR REQUIREMENTS OR EXPECTATIONS. THE CONTENT IS PROVIDED “AS IS” AND ON AN “AS AVAILABLE” BASIS. YOU EXPRESSLY AGREE THAT YOUR USE OF THE CONTENT IS AT YOUR SOLE RISK. PYMNTS SHALL HAVE NO LIABILITY FOR ANY INTERRUPTIONS IN THE CONTENT THAT IS PROVIDED AND DISCLAIMS ALL WARRANTIES WITH REGARD TO THE CONTENT, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, AND NON-INFRINGEMENT AND TITLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF CERTAIN WARRANTIES, AND, IN SUCH CASES, THE STATED EXCLUSIONS DO NOT APPLY. PYMNTS RESERVES THE RIGHT AND SHOULD NOT BE LIABLE SHOULD IT EXERCISE ITS RIGHT TO MODIFY, INTERRUPT, OR DISCONTINUE THE AVAILABILITY OF THE CONTENT OR ANY COMPONENT OF IT WITH OR WITHOUT NOTICE.

PYMNTS SHALL NOT BE LIABLE FOR ANY DAMAGES WHATSOEVER, AND, IN PARTICULAR, SHALL NOT BE LIABLE FOR ANY SPECIAL, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, OR DAMAGES FOR LOST PROFITS, LOSS OF REVENUE, OR LOSS OF USE, ARISING OUT OF OR RELATED TO THE CONTENT, WHETHER SUCH DAMAGES ARISE IN CONTRACT, NEGLIGENCE, TORT, UNDER STATUTE, IN EQUITY, AT LAW, OR OTHERWISE, EVEN IF PYMNTS HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

SOME JURISDICTIONS DO NOT ALLOW FOR THE LIMITATION OR EXCLUSION OF LIABILITY FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES, AND IN SUCH CASES SOME OF THE ABOVE LIMITATIONS DO NOT APPLY. THE ABOVE DISCLAIMERS AND LIMITATIONS ARE PROVIDED BY PYMNTS AND ITS PARENTS, AFFILIATED AND RELATED COMPANIES, CONTRACTORS, AND SPONSORS, AND EACH OF ITS RESPECTIVE DIRECTORS, OFFICERS, MEMBERS, EMPLOYEES, AGENTS, CONTENT COMPONENT PROVIDERS, LICENSORS, AND ADVISERS.

Components of the content original to and the compilation produced by PYMNTS is the property of PYMNTS and cannot be reproduced without its prior written permission.

The Generative AI Tracker® Series is a registered trademark of What’s Next Media & Analytics, LLC (“PYMNTS”).