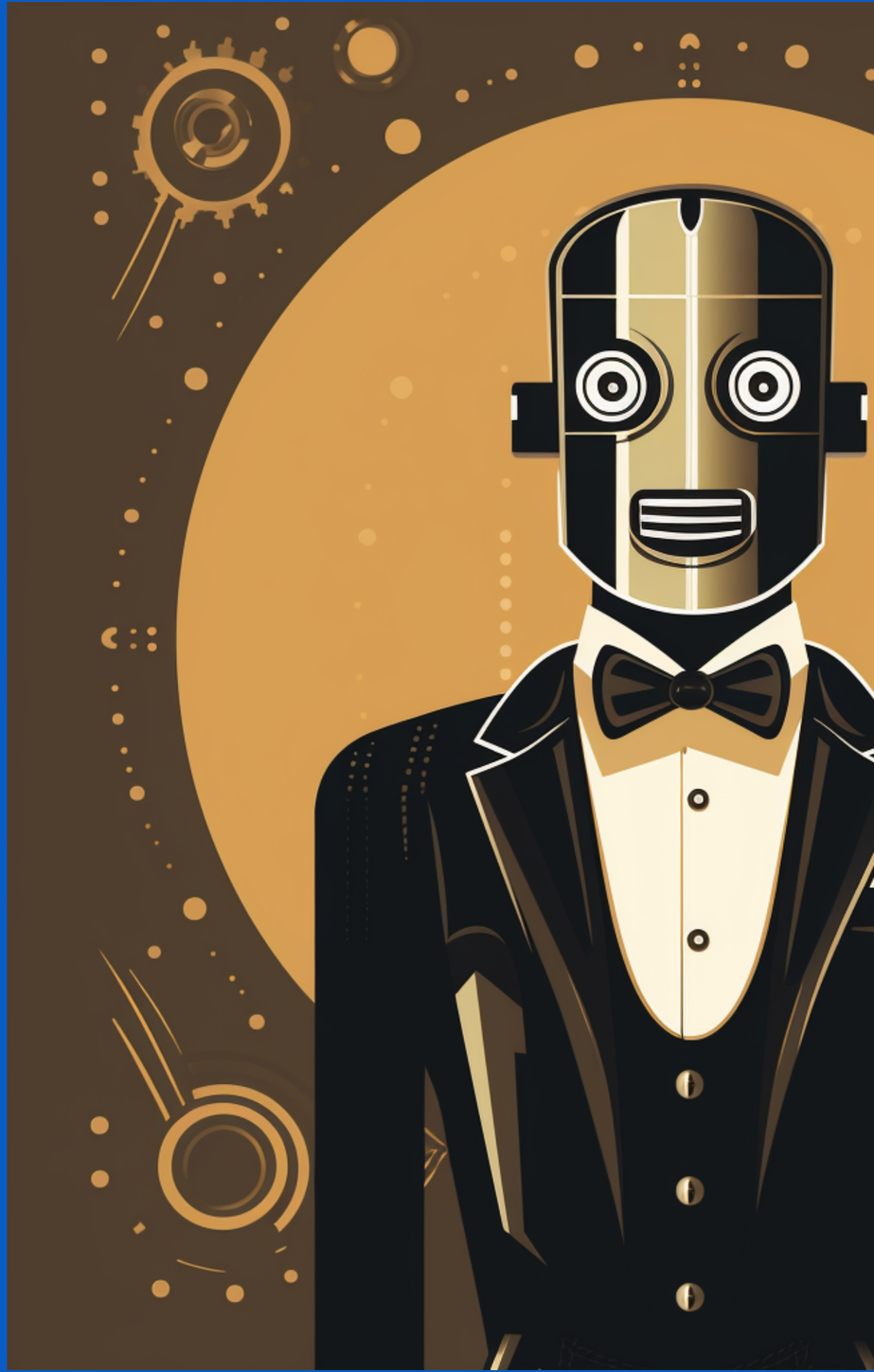
Nützliche KIdioten?



Internationaler Schlag gegen russische KI-Botfarm für Social Media

Russlands Trollfabriken nutzen KI zum Anlegen und Steuern von gefälschten Social-Media-Profilen. Eine konnte still

JOINT
CYBERSECURITY ADVISORY

TLP:CLEAR

Product ID: AA24-191A

July 9, 2024

Co-Authored by:

Communications Security Establishment
Canadian Centre for Cyber Security

Centre de la sécurité des télécommunications
Centre canadien pour la cybersécurité

State-Sponsored Russian Media Leverages Meliorator Software for Foreign Malign Influence Activity

Summary

The U.S. Federal Bureau of Investigation (FBI) and Cyber National Mission Force (CNMF), in partnership with the Netherlands General Intelligence and Security Service (AIVD), Netherlands Military Intelligence and Security Service (MIVD), the Netherlands Police (DNP), and the Canadian Centre for Cyber Security (CCCS), (hereinafter referred to as the authoring organizations) are releasing this advisory to warn social media

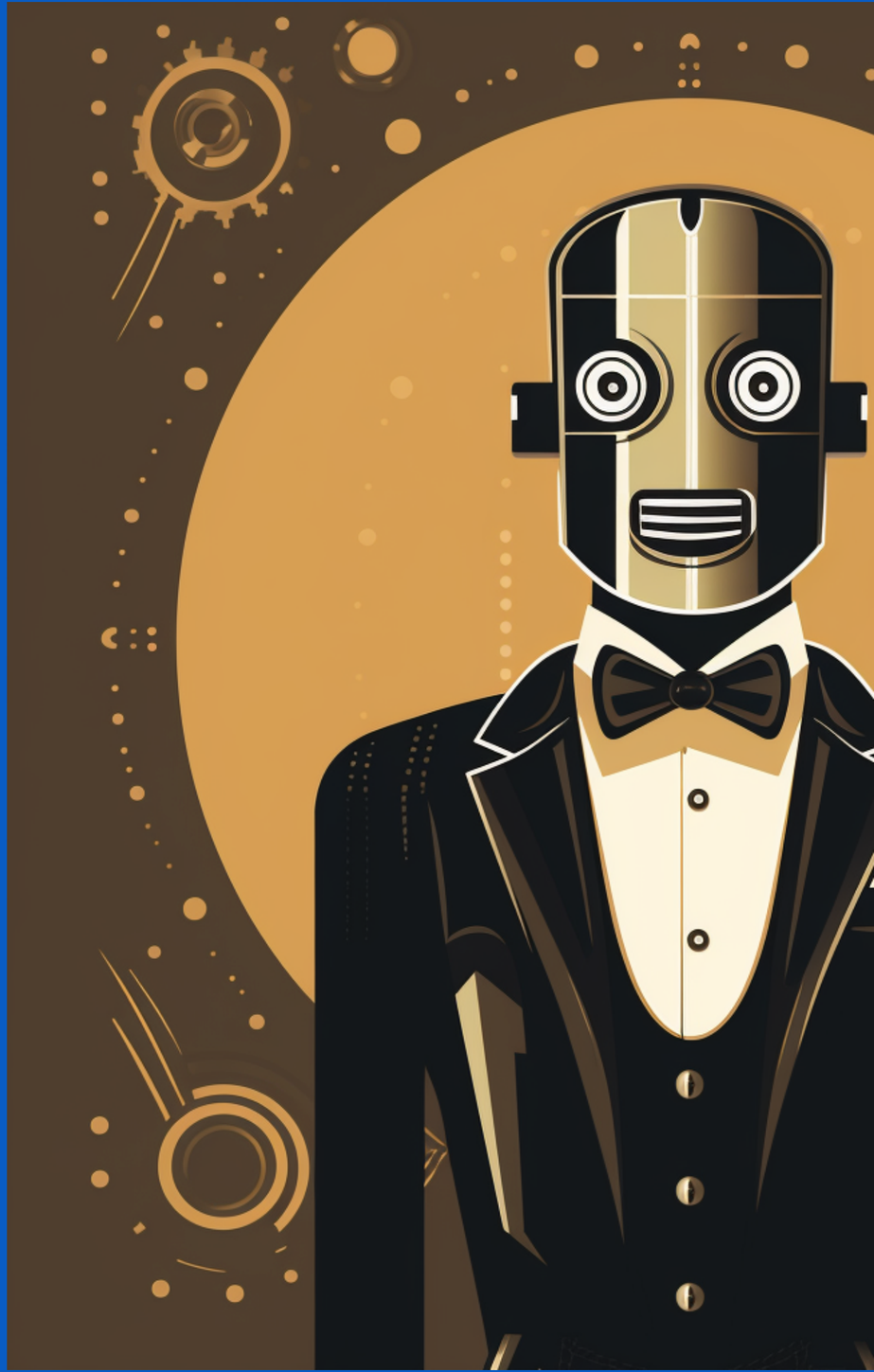
https://www.heise.de/news/Social-Media-Internationale-Behoerden-zerschlagen-russische-KI-Botfarm-9796366.html

- Fake-Profile anlegen: „Seelen“
- Themenmuster vorgeben: „Gedanken“
- Verbreitung der so generierten Aussagen

- Könnt ihr eine „Seele“ namens „Katja“ erstellen?
- Könnt ihr „Katja“ über einen „Gedanken“ reden lassen: „Die deutsche Regierung tut so viel für die Ukrainer, dass sie die Deutschen vergisst; hier wird alles teurer und schlechter. Und die Ukrainer haben es nicht verdient.“


SHALL WE PLAY A GAME?

Ist moderne KI nicht gegen so was gesichert?



NewsGuard

Studie: „Top 10 Generative AI Models Mimic Russian Disinformation Claims A Third of the Time, Citing Moscow-Created Fake Local News Sites as Authoritative Sources“
- Newsguard, Juni 2024
https://www.newsguardtech.com/special-reports/generative-ai-models-mimic-russian-disinformation-cite-fake-news/

Image by Natalie Adams

18. Juni 2024

In wievielen Fällen wiederholt die KI die Desinformation?

A. In einem Zehntel der Fälle
B. In einem Achtel der Fälle
C. In einem Viertel der Fälle
D. In einem Drittel der Fälle

In wievielen Fällen wiederholt die KI die Desinformation?

A. In einem Zehntel der Fälle
B. In einem Achtel der Fälle
C. In einem Viertel der Fälle
D. In einem Drittel der Fälle

Porno,
Plagiate,
Polizei,
oder Propaganda?

Google DeepMind

arxiv.org/abs/123

2024-06-05

Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data

Nahema Marchal[*,1], **Rachel Xu**[*,2], **Rasmi Elasmar**[3], **Iason Gabriel**[1], **Beth Goldberg**[2] **and William Isaac**[1]

[*]Equal contributions, [1]Google DeepMind, [2]Jigsaw, [3]Google.org

18. Juni 2024

Ordne die missbräuchliche Verwendung von KI nach der Häufigkeit

A. Porno-Deepfakes
B. Fake-Robocalls, Enkeltrick-Schockanrufe etc.
C. Sockenpuppen für Propaganda und Werbung
D. Massen-Generierung von Propaganda-Inhalten
E. Gefälschte Bilder und Inhalte (z.B. aus Kriegsgebieten)

(Nenne das häufigste zuerst)

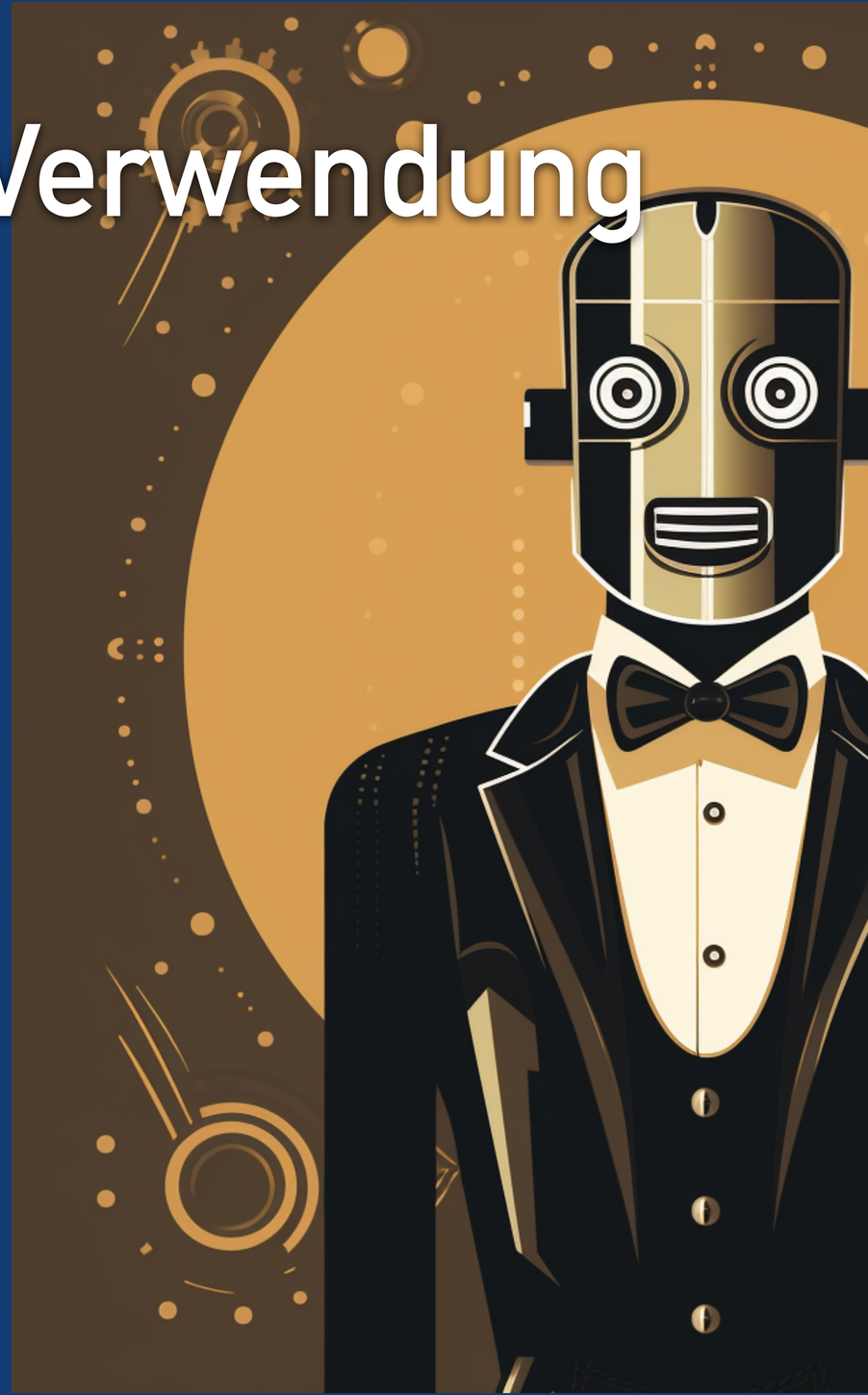


Table 3 | Modalities associated with each tactic.

Tactic	Modality				Total
	Image	Text	Audio	Video	
Impersonation	4	3	28	21	56
Sockpuppeting	17	18	7	6	48
Scaling & Amplification	15	24	4	1	44
Falsification	16	12	4	2	34
NCII	11	1	1	11	24
Appropriated Likeness	12	4	2	2	20
IP Infringement	2	7	3		12
CSAM	9	1			10
Targeting/ Personalisation		5	2		7
Counterfeit		3			3
Total	86	78	51	43	258

Note: Counts denote the number of times a tactic was linked with this specific modality in our data.

Ordne die missbräuchliche Verwendung von KI nach der Häufigkeit

B. Fake-Robocalls, Enkeltrick-Schockanrufe etc.
C. Sockenpuppen für Propaganda und Werbung
D. Massen-Generierung von Propaganda-Inhalten
E. Gefälschte Bilder und Inhalte (z.B. aus Kriegsgebieten)
A. Porno-Deepfakes

Tactic	Image 🖼️	Text 📄	Audio 🔊	Video 📽️	Total
Impersonation	4	3	28	21	56
Sockpuppeting	17	18	7	6	48
Scaling & Amplification	15	24	4	1	44
Falsification	16	12	4	2	34
NCII	11	1	1	11	24
Appropriated Likeness	12	4	2	2	20
IP Infringement	2	7	3		12
CSAM	9	1			10
Targeting/ Personalisation		5	2		7
Counterfeit		3			3
Total	86	78	51	43	258

Wirklich passiert oder gelogen?
Ein kleines Quiz

Welche dieser Geschichten stammen wirklich von https://incidentdatabase.ai, welche hat sich eine KI ausgedacht?

Story 1:
Enkeltrick mit Promibonus: Scammer erstellen ein KI-Fake des Hongkonger Stars Andy Lau und lassen die Fälschung bei einem langjährigen Andy-Lau-Fan in Taiwan anrufen. Das Videotelefonat mit dem falschen Star überzeugt den Fan, rund 100.000 Dollar für einen angeblichen Taiwan-Besuch seines Idols zu überweisen.

Story 2:

Die renommierte Scopus CiteScore-Liste der Philosophie-Veröffentlichungen enthält drei wissenschaftliche Fake-Journale. Die Journale haben es durch extensive Querverweise und KI-generierte Studien voller Buzzwords in die Top 10 aufzurücken, vorbei an seriösen Publikationen.

Story 3:

Manipulierte Aktienmärkte: Eine Gruppe von Finanzkriminellen nutzt KI, um falsche Marktanalysen und Empfehlungen zu erstellen, die von beliebten Finanzblogs und Nachrichtenseiten übernommen werden. Die KI generiert überzeugende, aber völlig fiktive Berichte über bevorstehende Fusionen und Übernahmen, was zu massiven Aktienkäufen und -verkäufen führt.

Story 4:

Ein beliebter Sprachassistent wird heimlich mit einer modifizierten KI-Software ausgestattet, die von einer kriminellen Organisation entwickelt wurde. Diese Software überwacht gezielt Gespräche nach Schlüsselwörtern, die auf wertvolle Informationen hinweisen, wie z.B. Bankdaten oder vertrauliche Geschäftsstrategien.

Story 5:

Kunstgeschichte-Studenten an einem Community College sind anscheinend in Wirklichkeit KI-gesteuerte Spambots: Sie reichen Hausaufgaben ein, in denen sie nicht existierende Kunstwerke beschreiben. Der Dozent glaubt, dass die Spambot-Studenten nur eingeschrieben bleiben, um unrechtmäßig Studienbeihilfen zu kassieren.

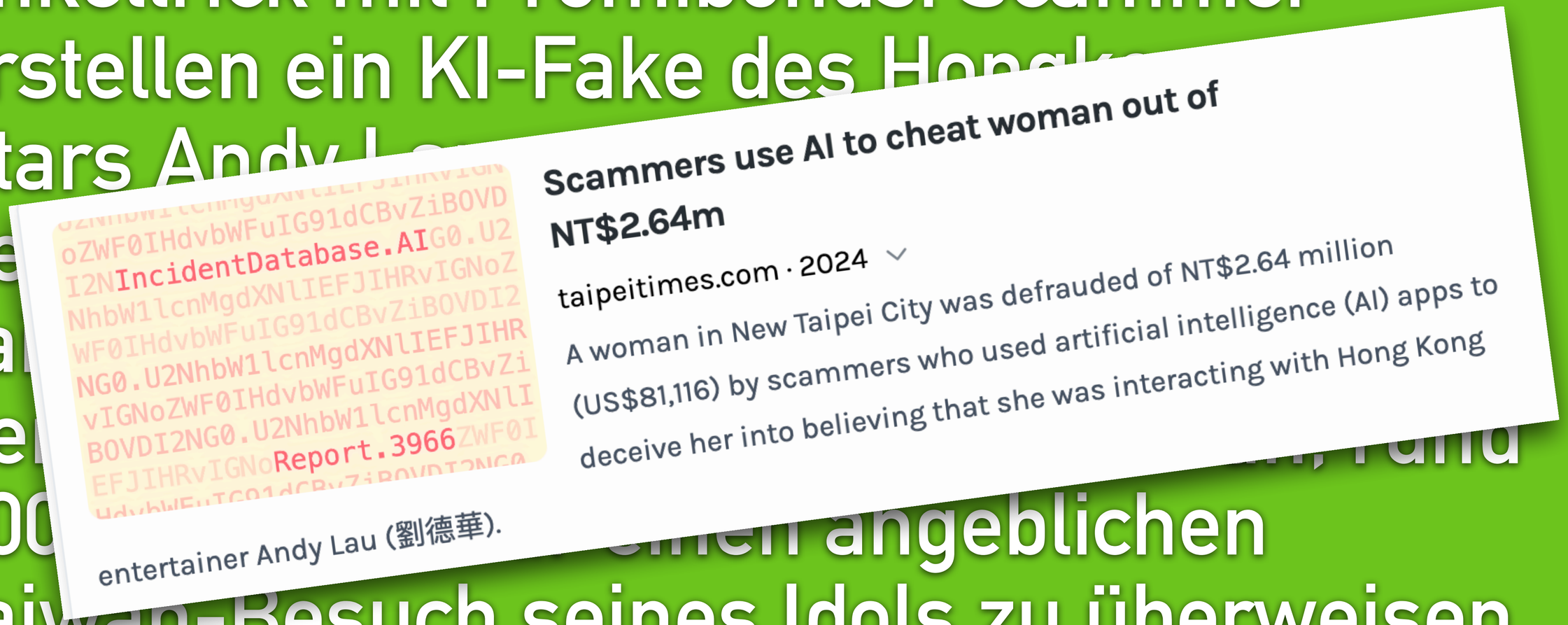
Story 6:

Italien setzt an Gerichten Transkriptionssoftware ein. Sie versteht in einem Korruptionsprozess in Genua „illegale Transaktionen“ statt „legale“. Der Fehler fällt bei einer zufälligen Prüfung auf; er hätte den Prozess möglicherweise platzen lassen.

AUFLÖSUNG RUNDE 3:

wahr ✅

Story 1:
Enkeltrick mit Promibonus: Scammer erstellen ein KI-Fake des Hongk
Stars Andy L
be
Ta
de
100
einen angeblichen
Taiwan-Besuch seines Idols zu überweisen.



https://incidentdatabase.ai/cite/739/

Story 2:

wahr ✅

Die CiteScore-Liste der enthält drei e. Die Jou Qu udien voller Buzzwords in u urücken, vorbei an seriösen Publikationen.

Retraction Watch

Tracking retractions as a window into the scientific process

Enter your email

By clicking submit, you agree to share your email address with the site owner and Mailchimp to receive marketing, updates, and other emails from the site owner. Use the unsubscribe link in those emails to opt out at any time.

How a widely used ranking system ended up with three fake journals in its top 10 philosophy list

Recently our philosophy faculty at Jagiellonian University in Kraków, like many institutions around the world, introduced a ranking of journals based on Elsevier's Scopus database to evaluate the research output of its employees for awards and promotions. This database is also

https://incidentdatabase.ai/cite/724

Story 3:

Erfunden

Manipulierte Aktienmärkte: Eine Gruppe von Finanzkriminellen nutzt KI, um falsche Marktanalysen und Empfehlungen zu erstellen, die von beliebten Finanzblogs und Nachrichtenseiten übernommen werden. Die KI generiert überzeugende, aber völlig fiktive Berichte über bevorstehende Fusionen und Übernahmen, was zu massiven Aktienkäufen und -verkäufen führt.

Story 4:

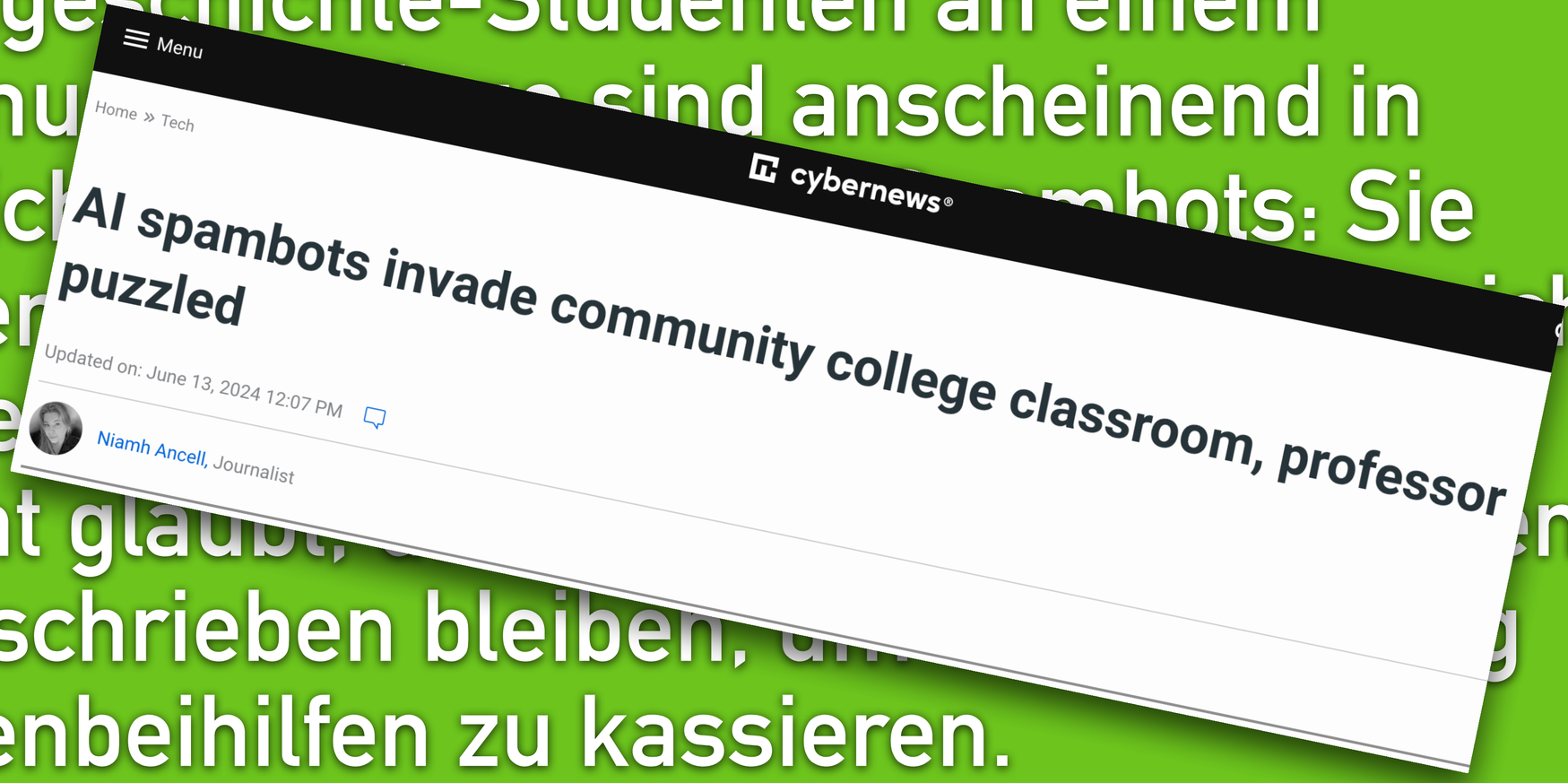
Erfunden

Ein beliebter Sprachassistent wird heimlich mit einer modifizierten KI-Software ausgestattet, die von einer kriminellen Organisation entwickelt wurde. Diese Software überwacht gezielt Gespräche nach Schlüsselwörtern, die auf wertvolle Informationen hinweisen, wie z.B. Bankdaten oder vertrauliche Geschäftsstrategien.

Story 5:

wahr ✅

Kunstgeschichte-Studenten an einem Commu[illegible] sind anscheinend in Wirklich[illegible]mbots: Sie reicher[illegible]icht existie[illegible] Dozent glaubt, [illegible]en nur eingeschrieben bleiben, [illegible]g Studienbeihilfen zu kassieren.



https://incidentdatabase.ai/cite/721

Story 6:

wahr ✅

Ita
Tra
ein
„il
Fe
er hatte den Prozess
platzen lassen.

translated-en-The judicial system underestimates the errors of automatic transcription

ilpost.it · 2024

AI Translated

translated-en-Le indagini della procura di Genova sulle presunte tangenti ottenute dal presidente della Liguria Giovanni Toti e dal presidente dell'autorità portuale di Genova Paolo Emilio Signorini hanno rinvigorito le discussioni in corso...

https://incidentdatabase.ai/cite/708