SoK: A Privacy Framework for Security Research Using Social Media Data

Kyle Beadle, Kieron Ivy Turk, Aliai Eusebi, Marilyne Ordekian, Enrico Mariconti, Yixin Zou, Marie Vasek

IEEE Symposium on Security and Privacy 2025

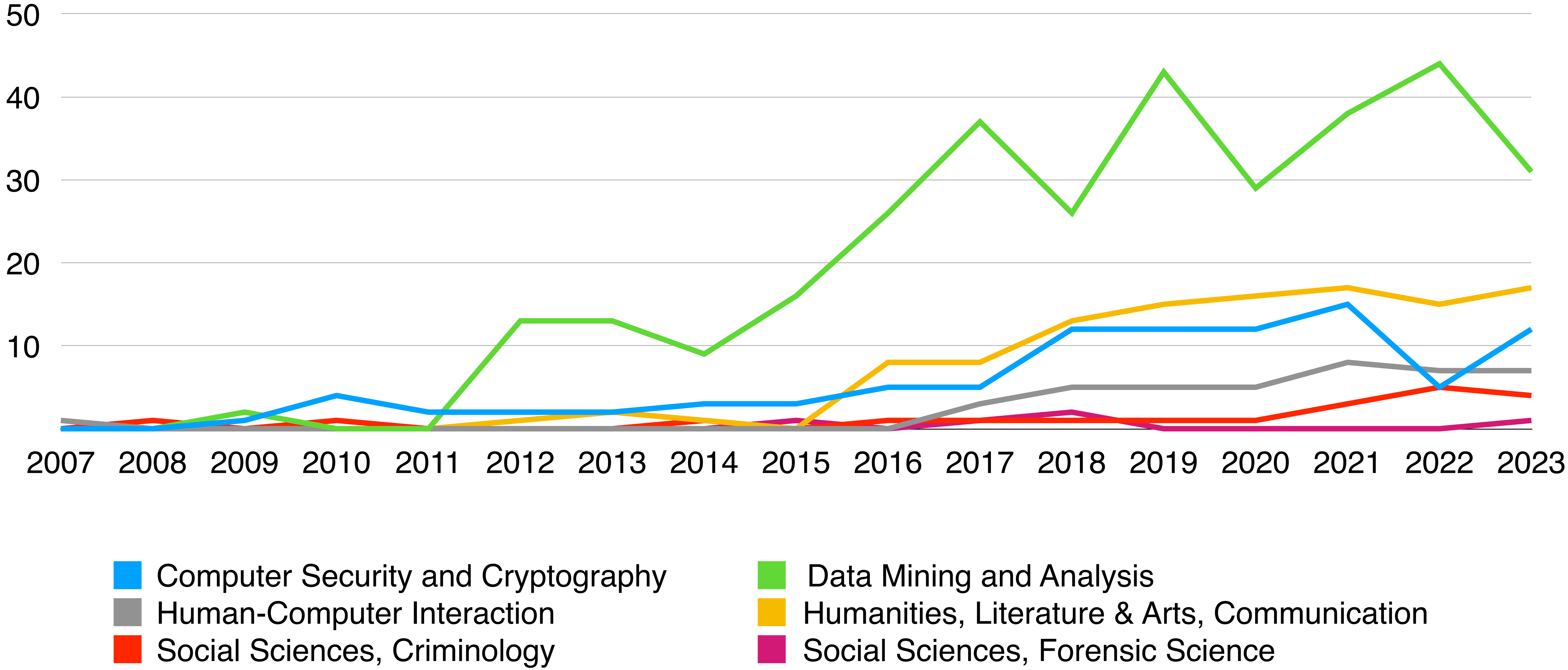
UCL

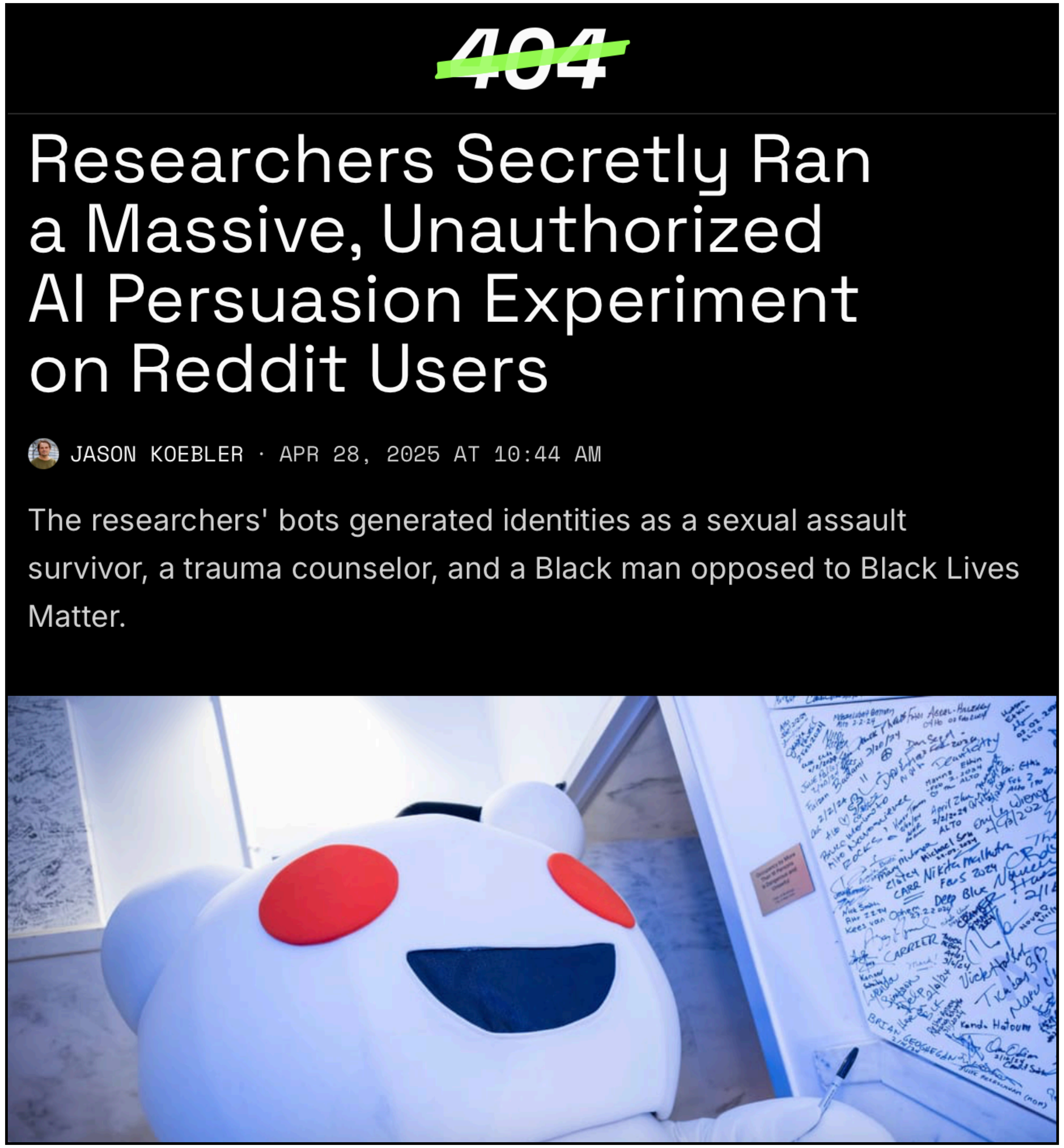
UNIVERSITY OF CAMBRIDGE

UKRI Engineering and Physical Sciences Research Council

MAX PLANCK INSTITUTE FOR SECURITY AND PRIVACY

Security and privacy research ❤️ social media data.





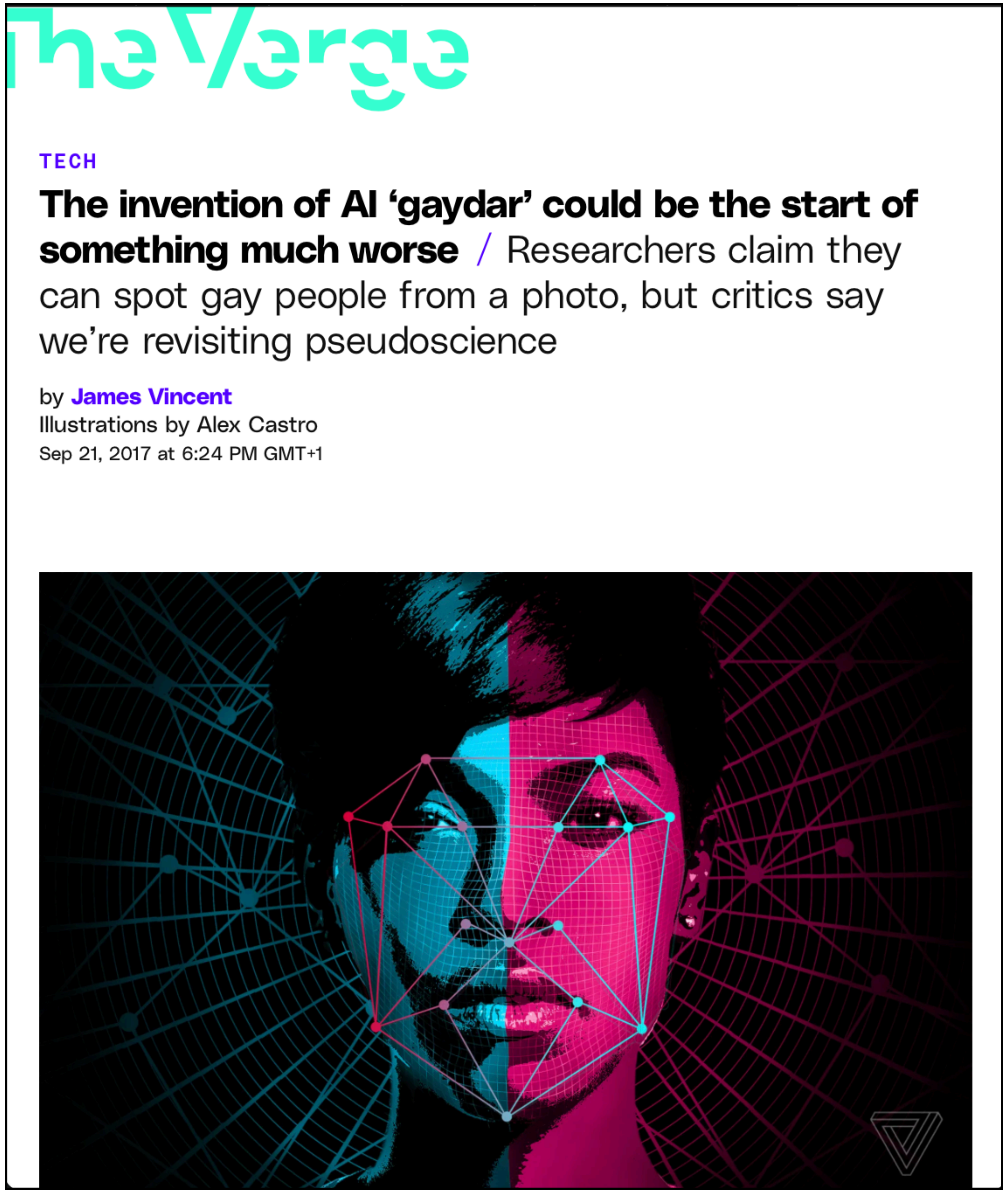
404

Researchers Secretly Ran a Massive, Unauthorized AI Persuasion Experiment on Reddit Users

JASON KOEBLER · APR 28, 2025 AT 10:44 AM

The researchers' bots generated identities as a sexual assault survivor, a trauma counselor, and a Black man opposed to Black Lives Matter.

Yet, data privacy is *often* an afterthought.



The Verge

TECH

The invention of AI 'gaydar' could be the start of something much worse / Researchers claim they can spot gay people from a photo, but critics say we're revisiting pseudoscience

by **James Vincent**
Illustrations by Alex Castro
Sep 21, 2017 at 6:24 PM GMT+1

Yet, data privacy is *often* an afterthought.

📋 Outline

1. How do security and privacy researchers handle privacy of social media data?
2. What privacy risks emerge from using social media data?
3. How do security and privacy researchers mitigate privacy risks?
4. Where do we go from here?

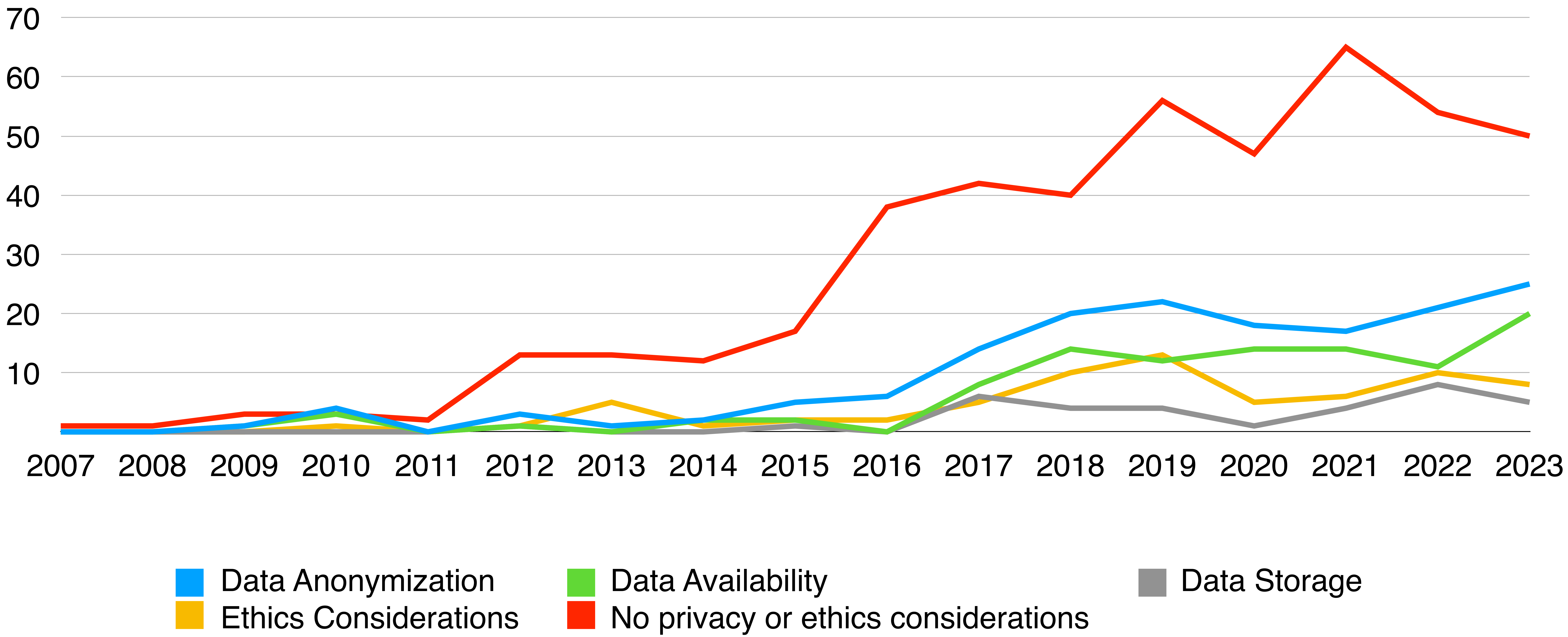
How do security and privacy researchers handle privacy of social media data?

📚 **601 papers** analyzed:

- **327** from Data Mining and Analysis (DMA)
- **113** from Humanities, Literature & Arts, Communication (HLAC)
- **96** from Computer Security and Cryptography (CSC)
- **40** from Human-Computer Interaction (HCI)
- **20** from Social Sciences, Criminology (SSC)
- **5** Social Sciences, Forensic Science (SSFS)

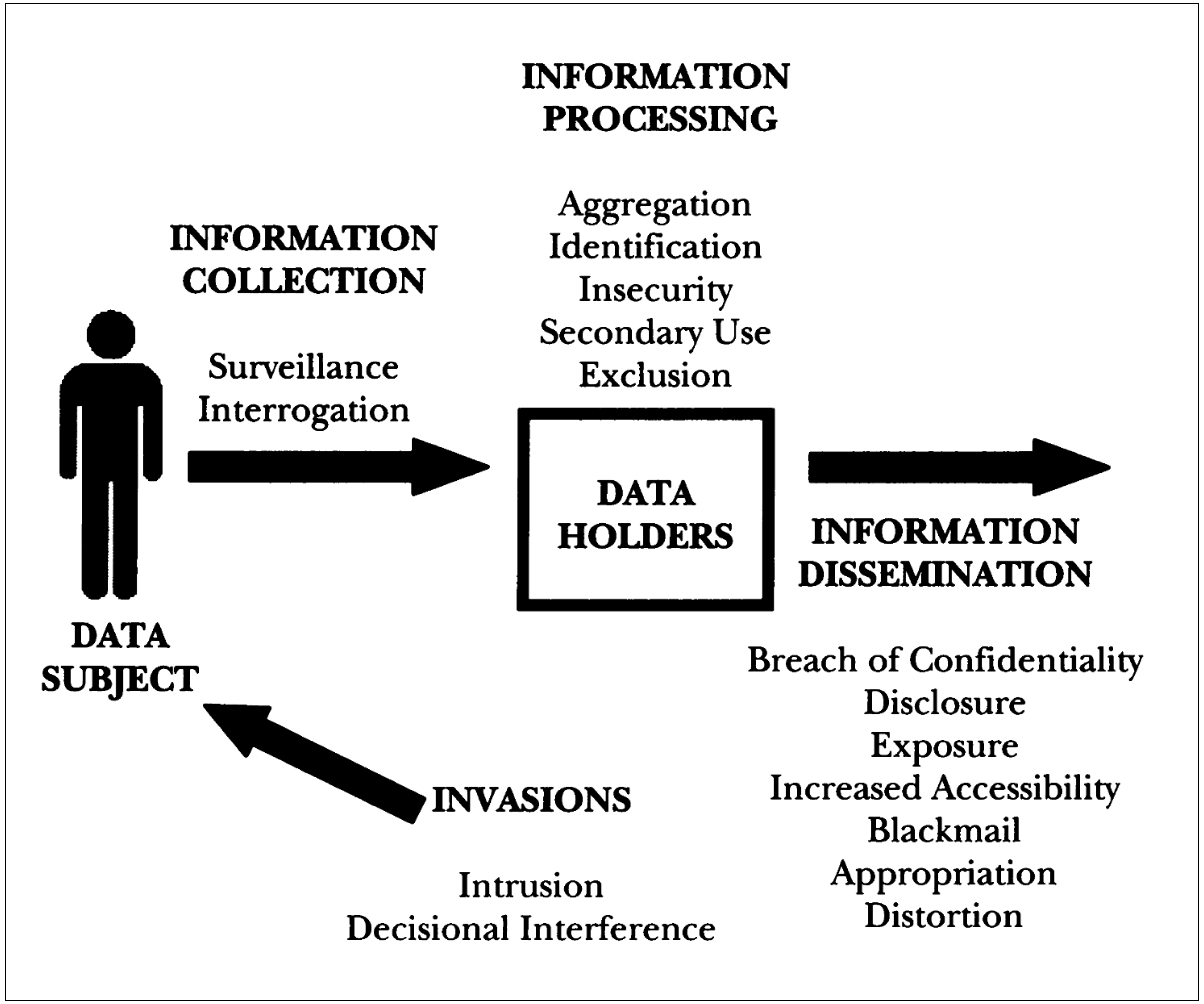
🚨 **Only 35% of security and privacy papers using social media data mention any considerations of data anonymization, availability, and storage.**

😞 Researchers are increasingly failing to report.



What privacy risks emerge from using social media data?

🔎 Solove's Taxonomy of Privacy



Daniel J Solove. A taxonomy of privacy. University of Pennsylvania Law Review, 154:477, 2005.

🔗 Adapting Solove’s Taxonomy

Process	Solove’s Taxonomy
Information Collection	Surveillance
Information Processing	Aggregation
	Identification
	Insecurity
	Exclusion
Information Dissemination	Disclosure
	Increased Accessibility
	Blackmail
	Distortion
Invasion	Intrusion
	Decisional Interference

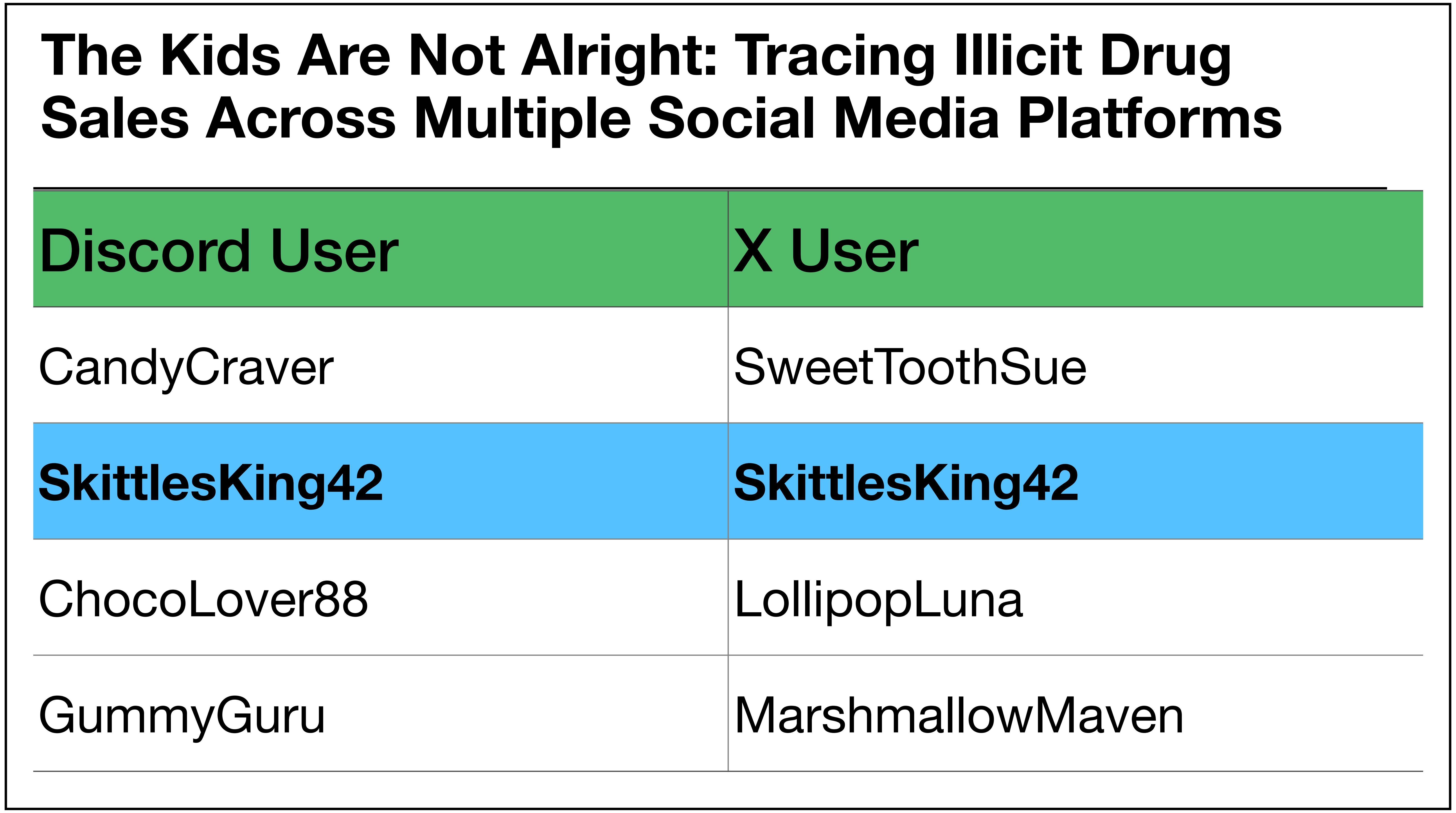
⚠️ Risk Manifestations

The Kids Are Not Alright: Tracing Illicit Drug Sales Across Multiple Social Media Platforms

- 1 million+ **Discord messages** collected from groups related to **university courses**
- 200k+ tweets from 20k+ users on **X** downloaded from a **public GitHub repo**

1. Identification
2. Increased Accessibility
3. Disclosure

⚠️ Risk Manifestations



The Kids Are Not Alright: Tracing Illicit Drug Sales Across Multiple Social Media Platforms

Discord User	X User
CandyCraver	SweetToothSue
SkittlesKing42	**SkittlesKing42**
ChocoLover88	LollipopLuna
GummyGuru	MarshmallowMaven

1. Identification
2. Increased Accessibility
3. Disclosure

⚠️ Risk Manifestations

The Kids Are Not Alright: Tracing Illicit Drug Sales Across Multiple Social Media Platforms

GitHub
OSF
kaggle

1. Identification
2. Increased Accessibility
3. Disclosure

⚠️ Risk Manifestations

The Kids Are Not Alright: Tracing Illicit Drug Sales Across Multiple Social Media Platforms

> “I never thought I'd be texting someone named 'SkittlesKing42' to get a bag of sour belts delivered to Smith Hall. 🤪🍬”

1. Identification
2. Increased Accessibility
3. Disclosure

How do security and privacy researchers mitigate privacy risks?

👷 Risk Mitigation

Certificate of Confidentiality

💼🗄️

Privacy Risk Analysis

🕵️🔒

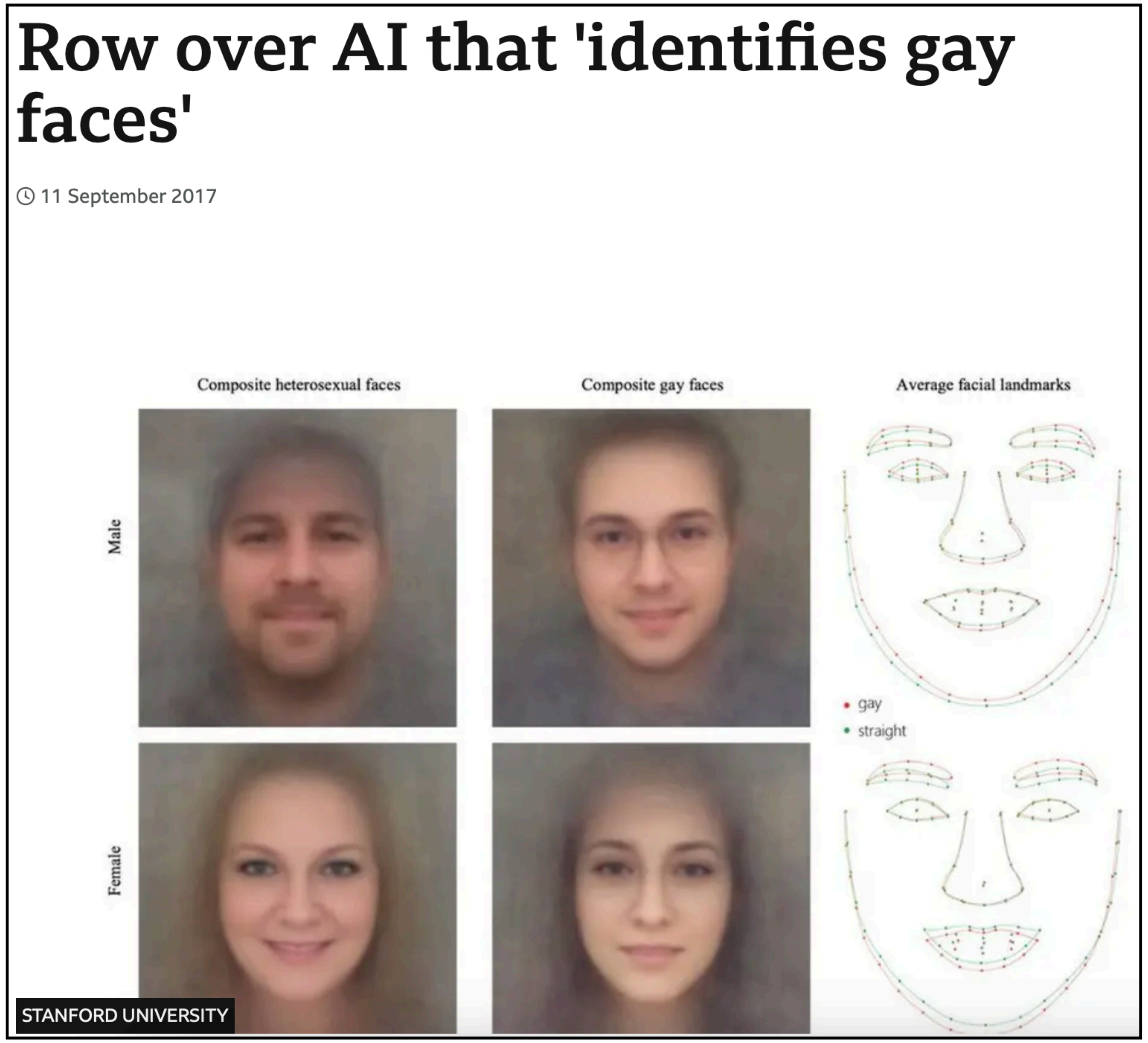
Data Donation

🫴👤

Where do we go from here?

📍 Implications

- Researchers disclosing risk
- Ethics boards/IRBs understanding risks
- Venues setting and enforcing expectations of social media data privacy



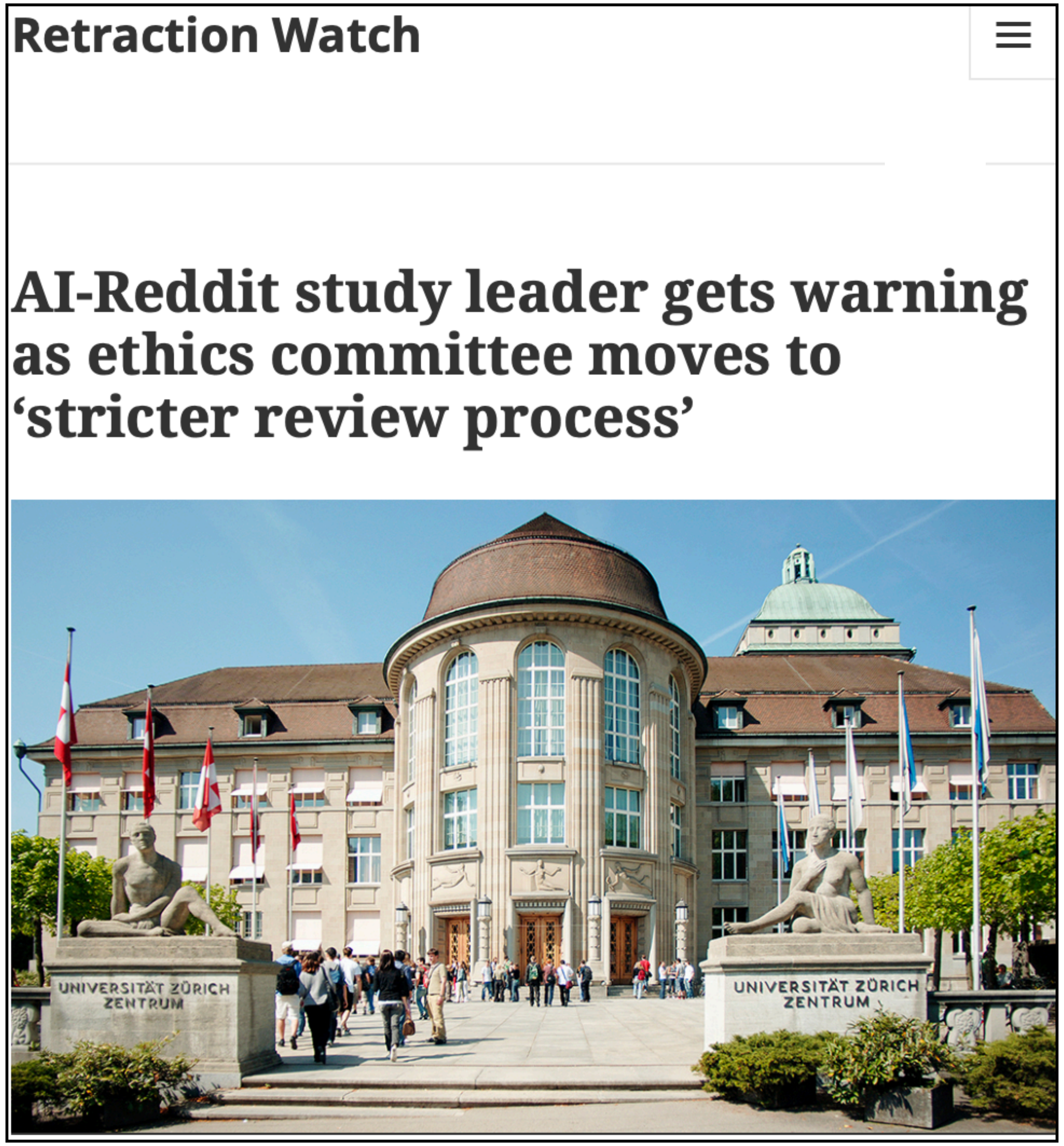
Row over AI that 'identifies gay faces'

11 September 2017

STANFORD UNIVERSITY

📍 Implications

- Researchers disclosing risk
- Ethics boards/IRBs understanding risks
- Venues setting and enforcing expectations of social media data privacy



Retraction Watch

AI-Reddit study leader gets warning as ethics committee moves to 'stricter review process'

📍 Implications

- Researchers disclosing risk
- Ethics boards/IRBs understanding risks
- Venues setting and enforcing expectations of social media data privacy

Ethical Considerations for Human Subjects Research

Submissions that describe experiments that could be viewed as involving human subjects, that analyze data derived from human subjects (even anonymized data), or that otherwise may put humans at risk should:

1. Disclose whether the research received an approval or waiver from each of the authors' institutional ethics review boards (IRB) if applicable.
2. Discuss steps taken to ensure that participants and others who might have been affected by an experiment were treated ethically and with respect.

If a submission deals with any kind of personal identifiable information (PII) or other kinds of sensitive data, the version of the paper submitted for review must discuss in detail the steps the authors have taken to mitigate harms to the persons identified. If a paper raises significant ethical and/or legal concerns, it will be checked by the REC and it might be rejected based on these concerns. The PC chairs will be happy to consult with authors about how this policy applies to their submissions.

💡 Key Takeaways

- Tools exist to respect user privacy—we must hold ourselves and each other accountable to implement them.
- Initiate privacy-conscious research design, not just compliance.
- Encourage documenting and reporting privacy decisions.

Paper link:

Kyle Beadle

kyle.beadle.22@ucl.ac.uk

https://kylebeadle.com

UCL

UNIVERSITY OF CAMBRIDGE

UKRI

Engineering and Physical Sciences Research Council

MAX PLANCK INSTITUTE FOR SECURITY AND PRIVACY