CSE127, Computer Security

Security + Privacy Beyond 127, Sociotechnical Security, The End

UC San Diego

Housekeeping

General course things to know

- PA5 due soon!
  - **3/14… good luck!**
- Final exam logistics!
  - **Final exam time:** Thursday, **3/19** at **8am**
  - **Final exam location:** Mosaic Lecture Hall 113 (has 250 seats so we don't need to be so cramped!)
  - **Final exam review session: 3/18** from 3:30pm – 5pm in CSE 1202
    - Can't be recorded, sorry if you have a conflict
  - Practice questions are online, solutions will be released this weekend

Final Exam Details

- Same format as the midterm: MCQ, SA, PA questions
- MCQ and SA are comprehensive over the entire class
    - My plan is to include ~5 – 10% of the midterm questions here; think of it like a “second chance.” **Looking over the midterm is a very good way to refresh yourself on the first half of the class!**
- PA questions will focus on PA4 and PA5
    - Best way to study for these is to refresh what you did (do) on PAs
    - Similarly, midterm will test your knowledge of PA material + extend a little bit
- One cheat sheet front and back is allowed

Today's Lecture

- We'll do a brief recap of all the places we've been, and give you a whirlwind taste of some of the stuff we didn't get to cover…

- We'll take 10 mins to fill out SETs

- I'll deep dive into my research area — sociotechnical cybersecurity — and a few recent projects I've really enjoyed working on

- None of this is on the final, so don't worry about notes, but because I'm me I'm still going to ask you to participate

CSE127, in sum

Computer security is all about trust

- If you were to take one thing away from this class, it’s that **trust is a first-class citizen** in all of computer security
  - Who do I trust, and why do I trust them to do X…. this is the fundamental question you must always ask yourself in security (and probably in life too)
- You can chart out *trust* in almost every single topic area we discussed!

Trust issues

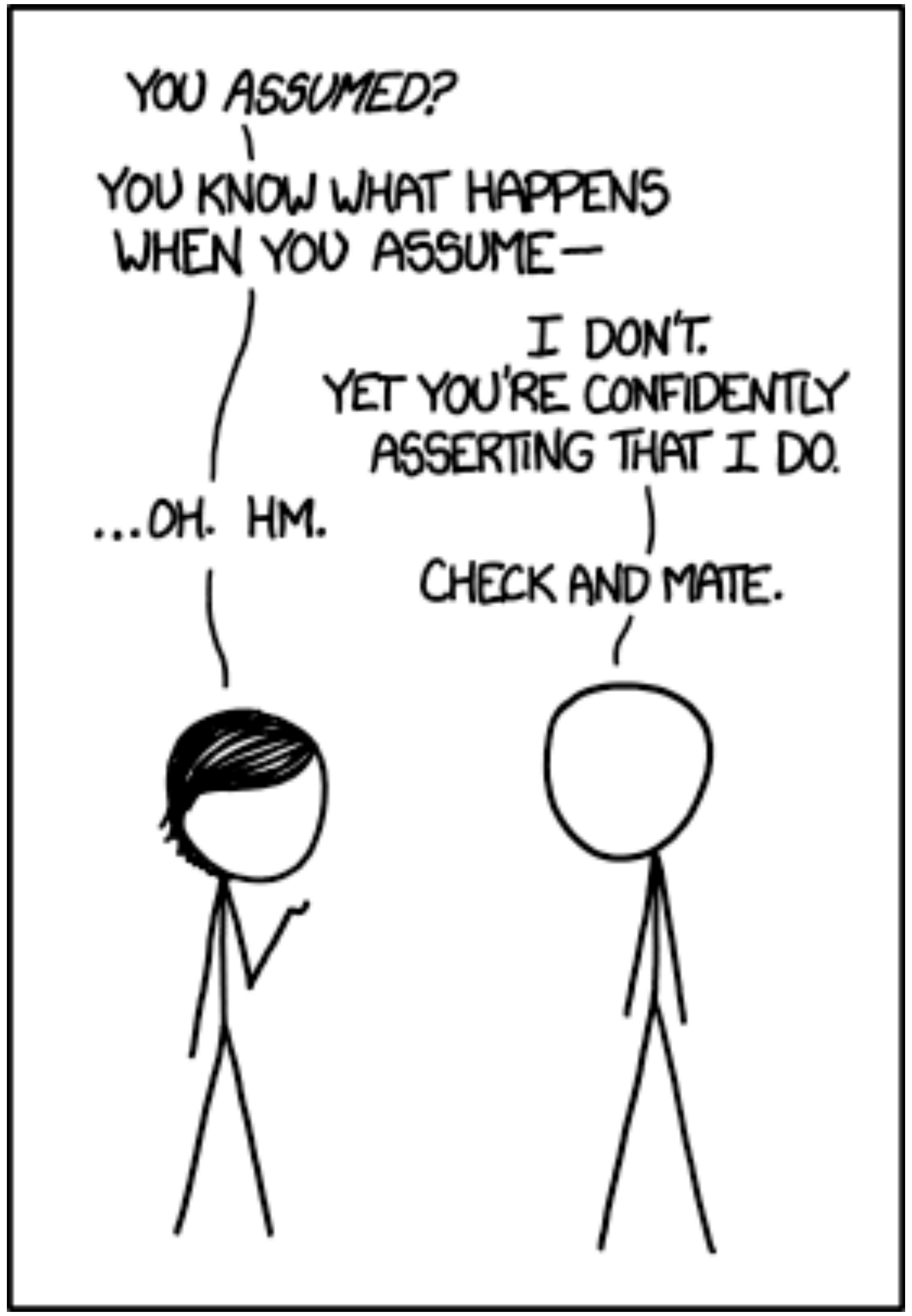
- AppSec
  - I trust the *C runtime* to not execute code that I, the developer, didn't put there. Why? How would the runtime even know?
  - Solution: Don't let code run from anywhere someone else could potentially be putting code! (D^X)… **trust the user and CPU less.**
- Systems Sec
  - I trust the *CPU* to not leak

Trust issues

- Identify three “trust issues” we’ve talked about or you’ve experienced in class. What was the trust issue? Why did the issue arise? Where might that have gone wrong? How did we fix it?

Computer security is all about assumptions

- You know what they say about assumptions…
- The attackers job is to **interrogate the assumptions made by the developers** to break the system
- The defenders job is to **enumerate the assumptions they are making** and ensure proper protections don't break invariants
    - This is the cat and mouse struggle


YOU ASSUMED?
YOU KNOW WHAT HAPPENS WHEN YOU ASSUME—
I DON'T. YET YOU'RE CONFIDENTLY ASSERTING THAT I DO.
...OH. HM.
CHECK AND MATE.

https://xkcd.com/1339/

Assuming things

- NetSec
  - Architecturally, we *assume* that because we're getting a packet from an IP, it must've come from that IP….
    - Obviously, this is **wrong**. No authentication on packets, so anyone can spoof an IP (or a DNS entry…. or an SMTP packet… or anything…)
- WebSec
  - We *assume* that the underlying SQL engine won't execute arbitrary deletion requests. Why? No reason!
  - We *assume* users will not try to manipulate the underlying SQL query generation. Why? No reason!

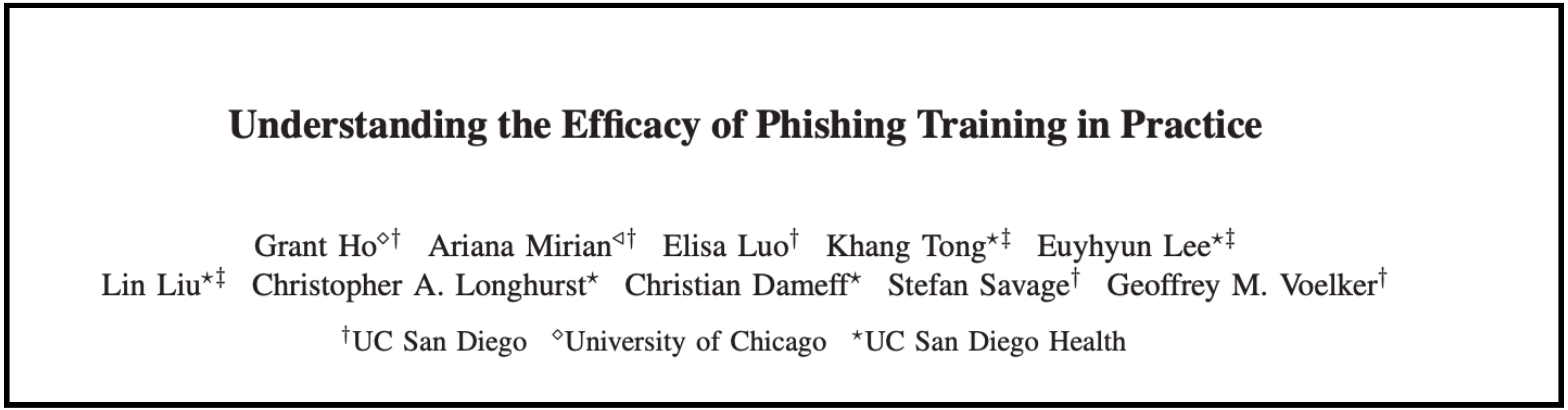
Assumptions in the real world — security education

Free Phishing Awareness Training
FREE
keepnet

PHISHING AWARENESS

What are the assumptions being made by the creators of this training?

Assumptions in the real world — security education

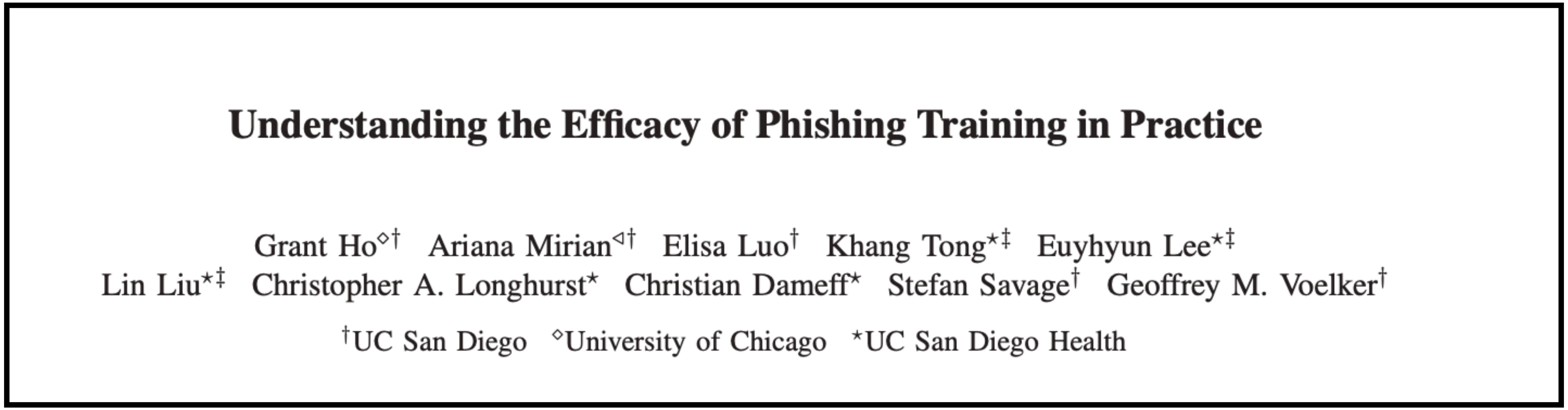


Understanding the Efficacy of Phishing Training in Practice

Grant Ho[◇†] Ariana Mirian[◁†] Elisa Luo[†] Khang Tong[⋆‡] Euyhyun Lee[⋆‡]
Lin Liu[⋆‡] Christopher A. Longhurst[⋆] Christian Dameff[⋆] Stefan Savage[†] Geoffrey M. Voelker[†]

[†] UC San Diego [◇] University of Chicago [⋆] UC San Diego Health

Assumptions in the real world — security education



Understanding the Efficacy of Phishing Training in Practice

Grant Ho[◇†] Ariana Mirian[◁†] Elisa Luo[†] Khang Tong[⋆‡] Euyhyun Lee[⋆‡]
Lin Liu[⋆‡] Christopher A. Longhurst[⋆] Christian Dameff[⋆] Stefan Savage[†] Geoffrey M. Voelker[†]

[†]UC San Diego [◇]University of Chicago [⋆]UC San Diego Health

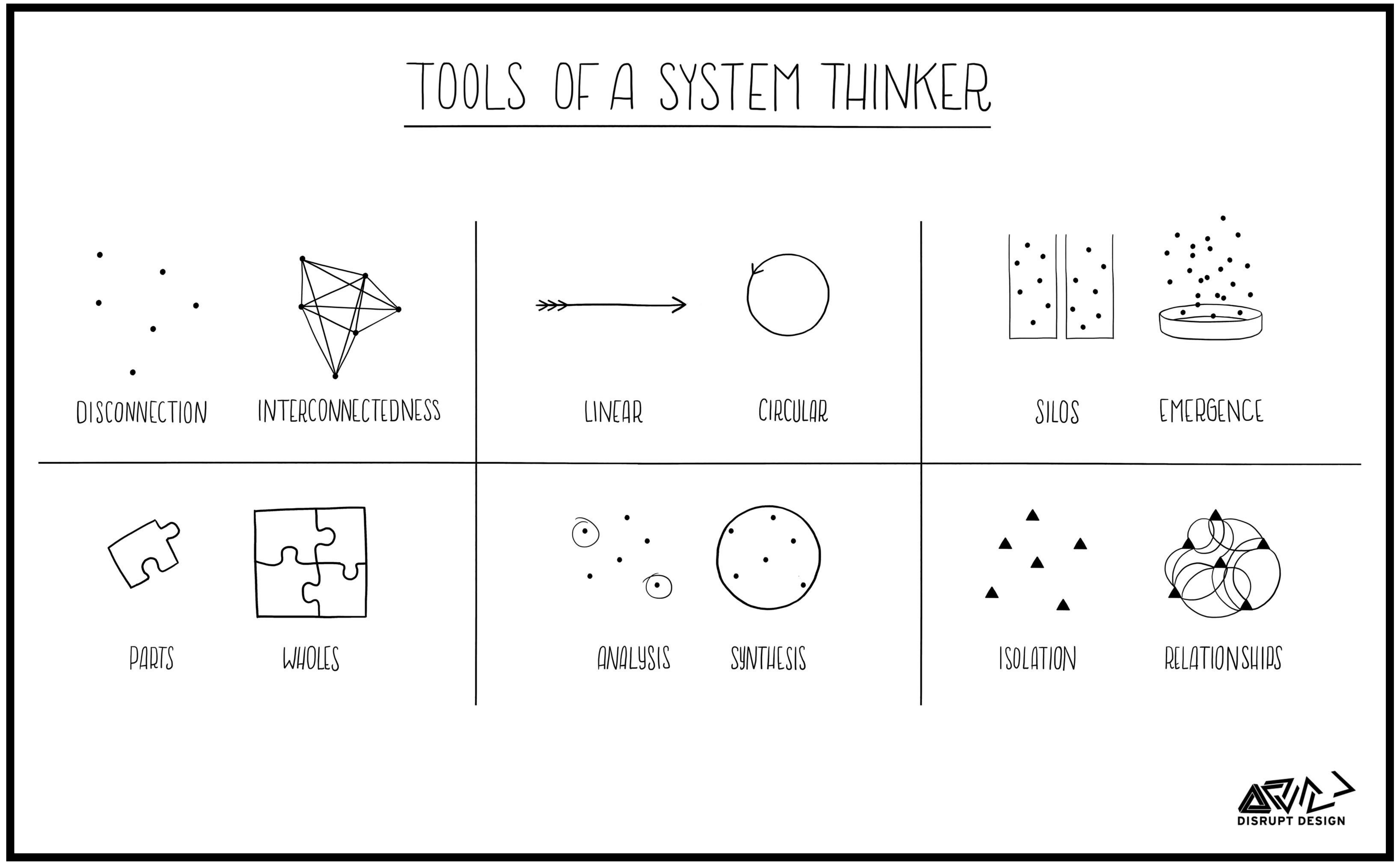
Taken together, our results suggest that anti-phishing training programs, in their current and commonly deployed forms, are unlikely to offer significant practical value in reducing phishing risks.

Computer security is about *system thinking*

- "System thinking is simply thinking about something a *system* — the existence of entities, the parts, the chunks, the pieces, and the relationships between them." – Edward Crawley, MIT

- Computer scientists like to think about the world in units and chunks
  - But security people like to think about the whole picture

- Systems include: software, hardware, and *people*….



Consider this scenario

- You’ve surreptitiously stolen someone’s Bitcoin hardware wallet, but it’s password protected. What are the systems in play that you might exploit to get access?

Consider this scenario

- You've surreptitiously stolen someone's Bitcoin hardware wallet, but it's password protected. <u>What are the systems in play that you might exploit to get access?</u>

    - Hack the hardware (open it up, cache timing side channel, etc.)

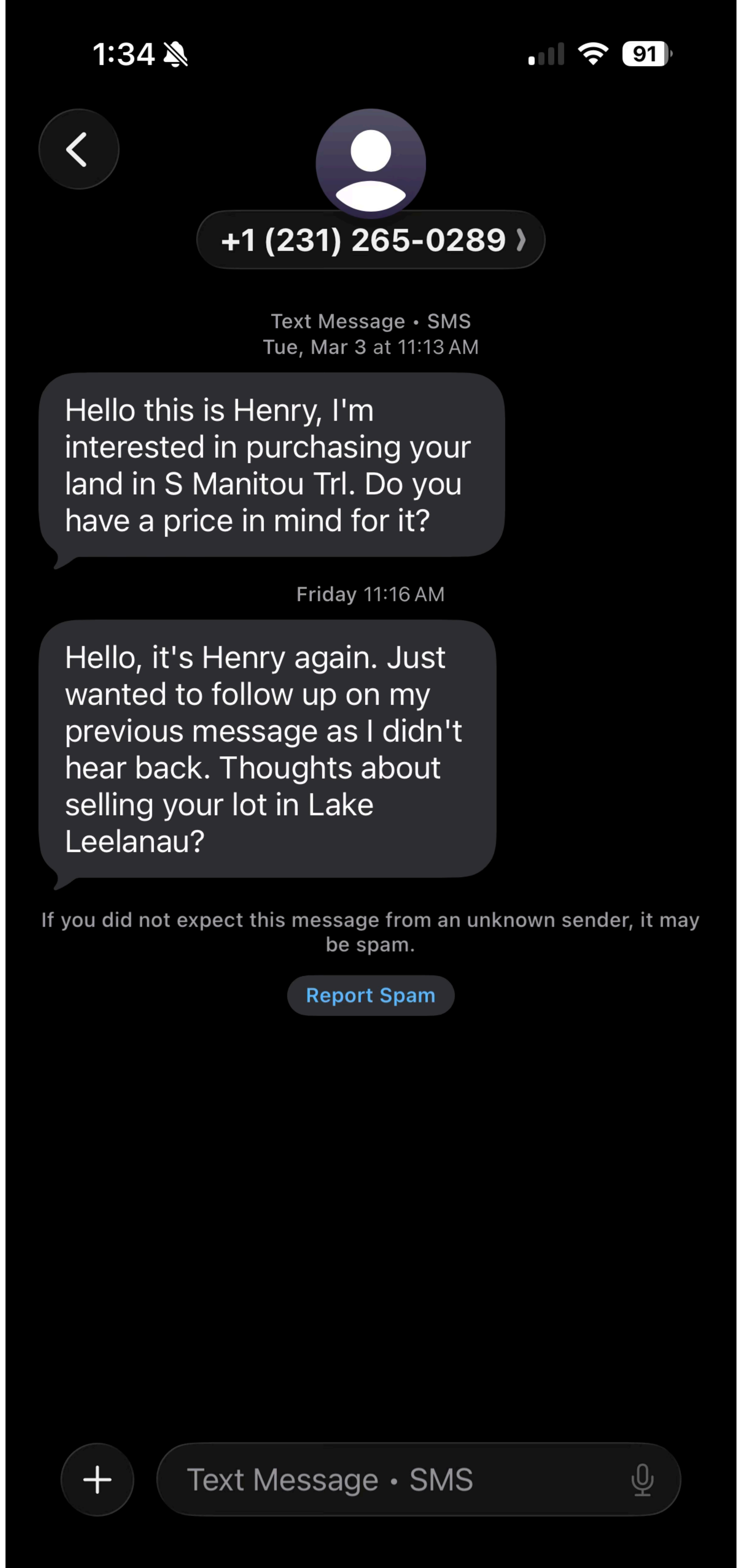
    - Hack the software (connect it to a machine, see any bugs in the interface)

Consider this scenario

- You’ve surreptitiously stolen someone’s Bitcoin hardware wallet, but it’s password protected. What are the systems in play that you might exploit to get access?
  - Hack the hardware (open it up, cache timing side channel, etc.)
  - Hack the software (connect it to a machine, see any bugs in the interface)
  - Hack the person!

Pig-Butchering Scams

- Fraudsters gain *trust* over time and deceive user into divulging **secret information** or even *sending money*
  - "Pig-Butchering" — fattening up the pig before… you know
- This is **extremely common**
- Ongoing research — study these types of scams using a VictimLLM honeypot…
  - Essentially trick fraudsters into think they're building trust, but it's really with a bot, and we're controlling the bot



Text Message • SMS
Tue, Mar 3 at 11:13 AM

Hello this is Henry, I'm interested in purchasing your land in S Manitou Trl. Do you have a price in mind for it?

Friday 11:16 AM

Hello, it's Henry again. Just wanted to follow up on my previous message as I didn't hear back. Thoughts about selling your lot in Lake Leelanau?

If you did not expect this message from an unknown sender, it may be spam.

Three big things to consider in security

- **Trust:** Who do I trust, and why?
- **Assumption:** What assumptions am I making about what I’m interacting with, and why?
- **System:** Can I consider the ways each component *interconnects*, and what assumptions (trust or otherwise) are being made about those interconnects that can be exploited?

We covered a lot of ground…

- Application security
- Systems security
- Web security
- Network security
- Cryptography

But we also barely scratched the surface!

- Usable Security and Privacy
- LLM / AI security
- Blockchain
- Privacy (broadly)
- Embedded device and hardware security
- Zero knowledge and multi-party computation

- Authentication (passwords, 2FA, biometrics)
- Mobile security
- Fraud, Malware, Spam, Crime
- Economics of cybersecurity
- Policy
- Cyberwar
- Sociotechnical security

But we also barely scratched the surface!

- Usable Security and Privacy
- LLM / AI security
- Blockchain
- Privacy (broadly)
- Embedded device and hardware security
- Zero knowledge and multi-party computation

- **Authentication (passwords, 2FA, biometrics)**
- Mobile security
- Fraud, Malware, Spam, Crime
- **Economics of cybersecurity**
- Policy
- **Cyberwar**
- **Sociotechnical security**

Authentication

- This whole time, we’re talking about computers… but people also matter!
- **Huge** part of security that you interact with every single day. Asks a fundamental question: **what is a human?**
- How do you authenticate yourself in typical computer systems?

How do we authenticate people…?

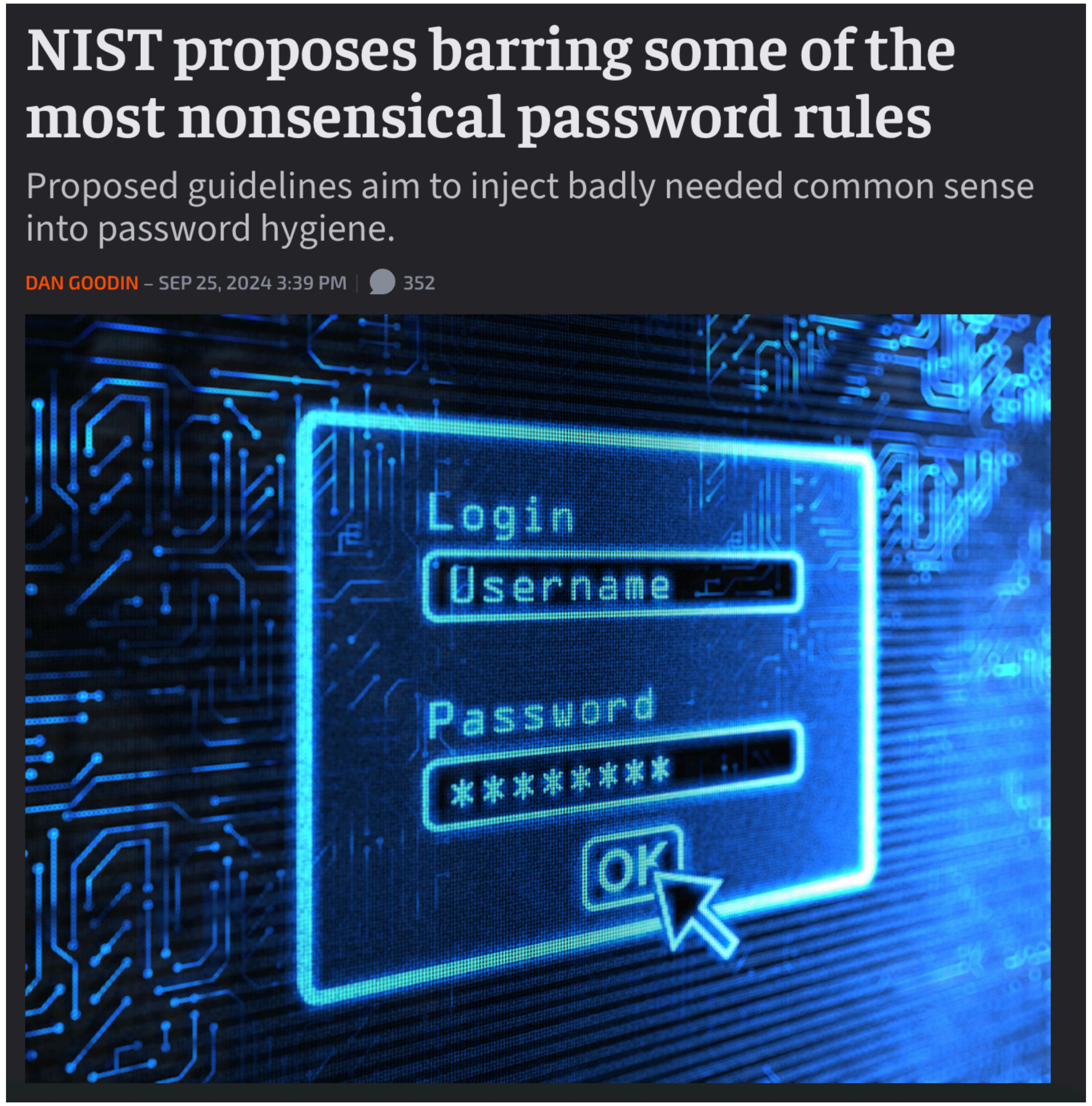
- This whole time, we’re talking about computers… but people also matter!
- **Huge** part of security that you interact with every single day. Asks a fundamental question: **what is a human?**
- How do you authenticate yourself in typical computer systems?
    - Passwords
    - Patterns (Android)
    - Private key
    - Passkeys (new)
    - Name, PID —> that’s how you authenticate yourself on the final exam

Sadly, people are bad at passwords

- How many of you have ever re-used the same password on multiple services? (Don’t answer that)
- How many of you, when trying to make an account as quickly as possible and meet the “alphanumeric+symbol” requirement, appended “123!” (Don’t answer that)
- How many of you keep up with which of your passwords have **already been breached?** (You can answer that)

Sadly, we make passwords really hard!

- National Institute for Standards and Technology produces password guidance year after year…
  - Character complexity
  - Length
  - PW rotation policies
- You name it….

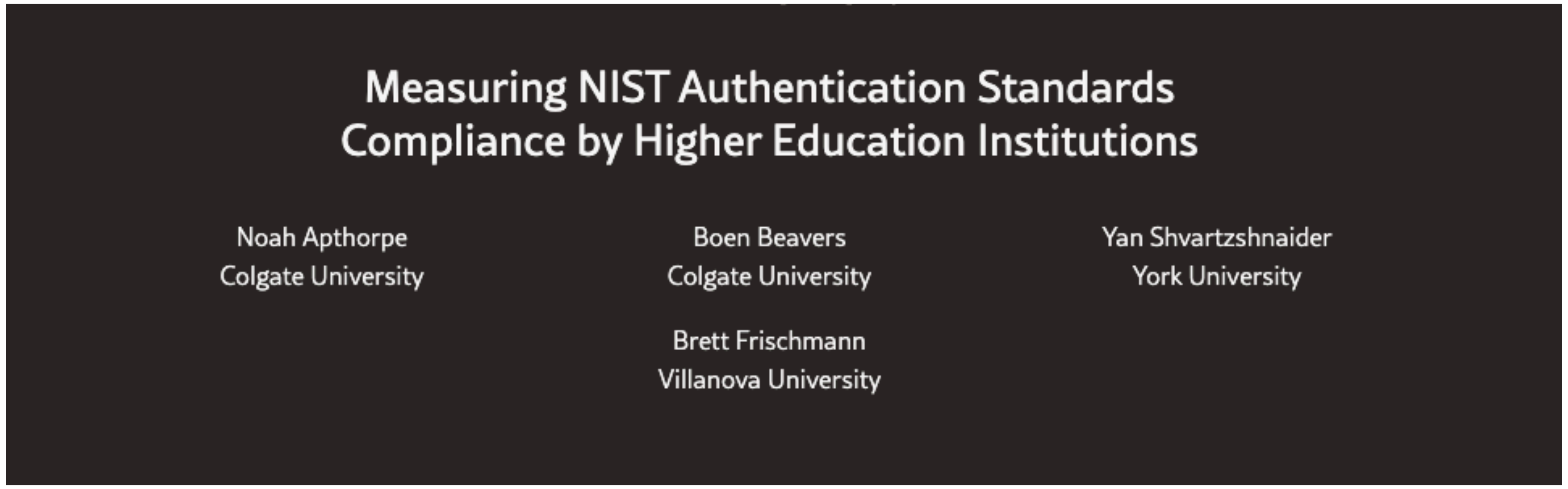


NIST proposes barring some of the most nonsensical password rules

Proposed guidelines aim to inject badly needed common sense into password hygiene.

DAN GOODIN – SEP 25, 2024 3:39 PM | 352

Sadly, we make passwords really hard!



Measuring NIST Authentication Standards Compliance by Higher Education Institutions

Noah Apthorpe
Colgate University

Boen Beavers
Colgate University

Yan Shvartzshnaider
York University

Brett Frischmann
Villanova University

We also find widespread noncompliance with standards for password expiration, password composition rules, and knowledge-based authentication… expert cybersecurity recommendations are not effectively reaching practitioners.

You really should be using a password manager

1Password

bitwarden

DASHLANE

Multifactor Authentication

- Passwords get compromised…. so how do you protect yourself?
- Three types of authentication factors:
  - ***Something you know***
  - ***Something you have***
  - ***Something you are***
- When we combine these together, we get *multi-factor authentication* (best defense against bad actors)
- **You should have MFA on every single high value account you own.**

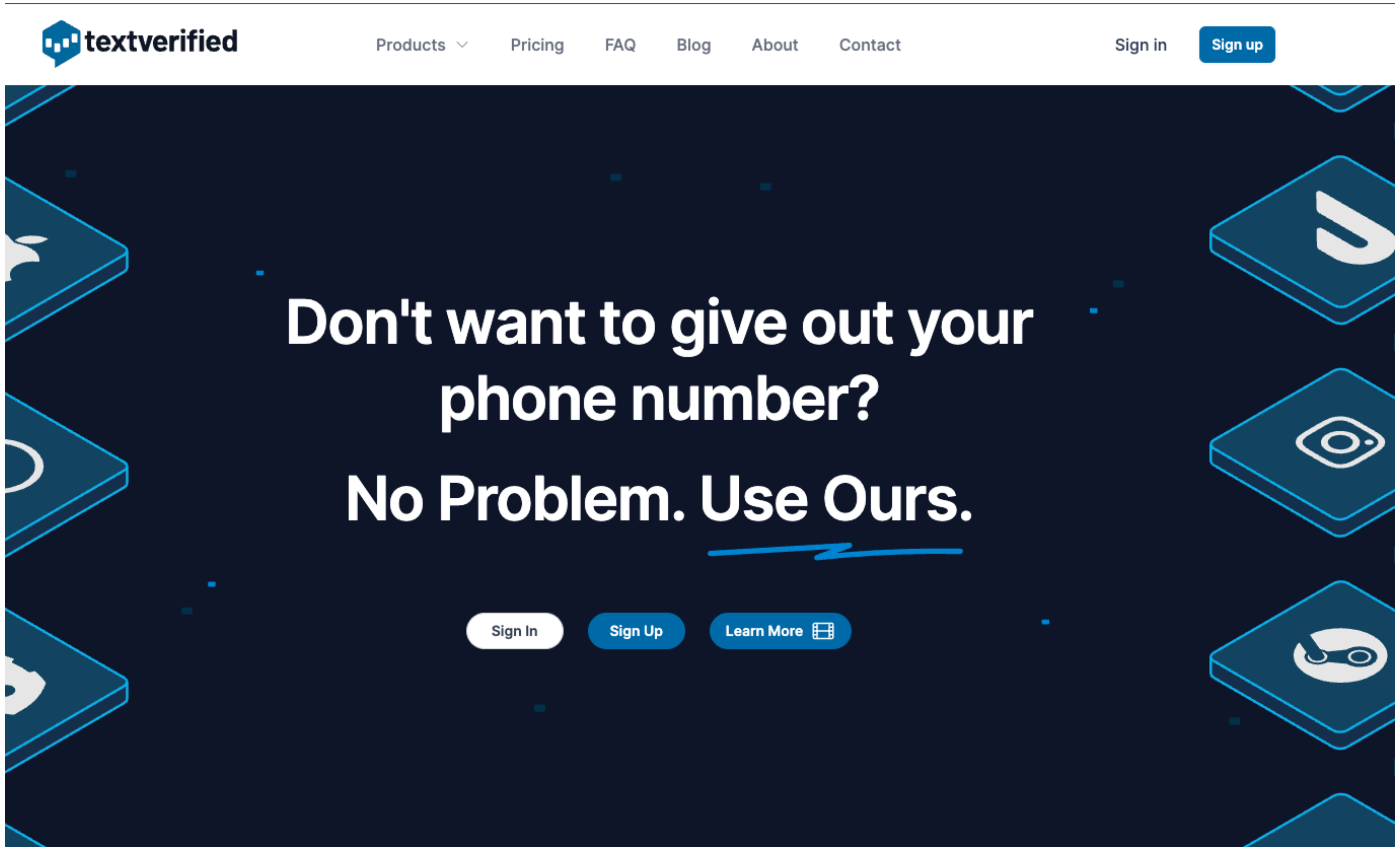
Remember, all of these can be bypassed individually…

Economics of Security + Privacy

- Let’s play a game…

Economics of Security

- Everything has a price!
    - CPUs, memory, IP address, bandwidth, storage, account credentials, etc.
    - Phone numbers!
- Attackers constantly seeking to monetize your compromised resources

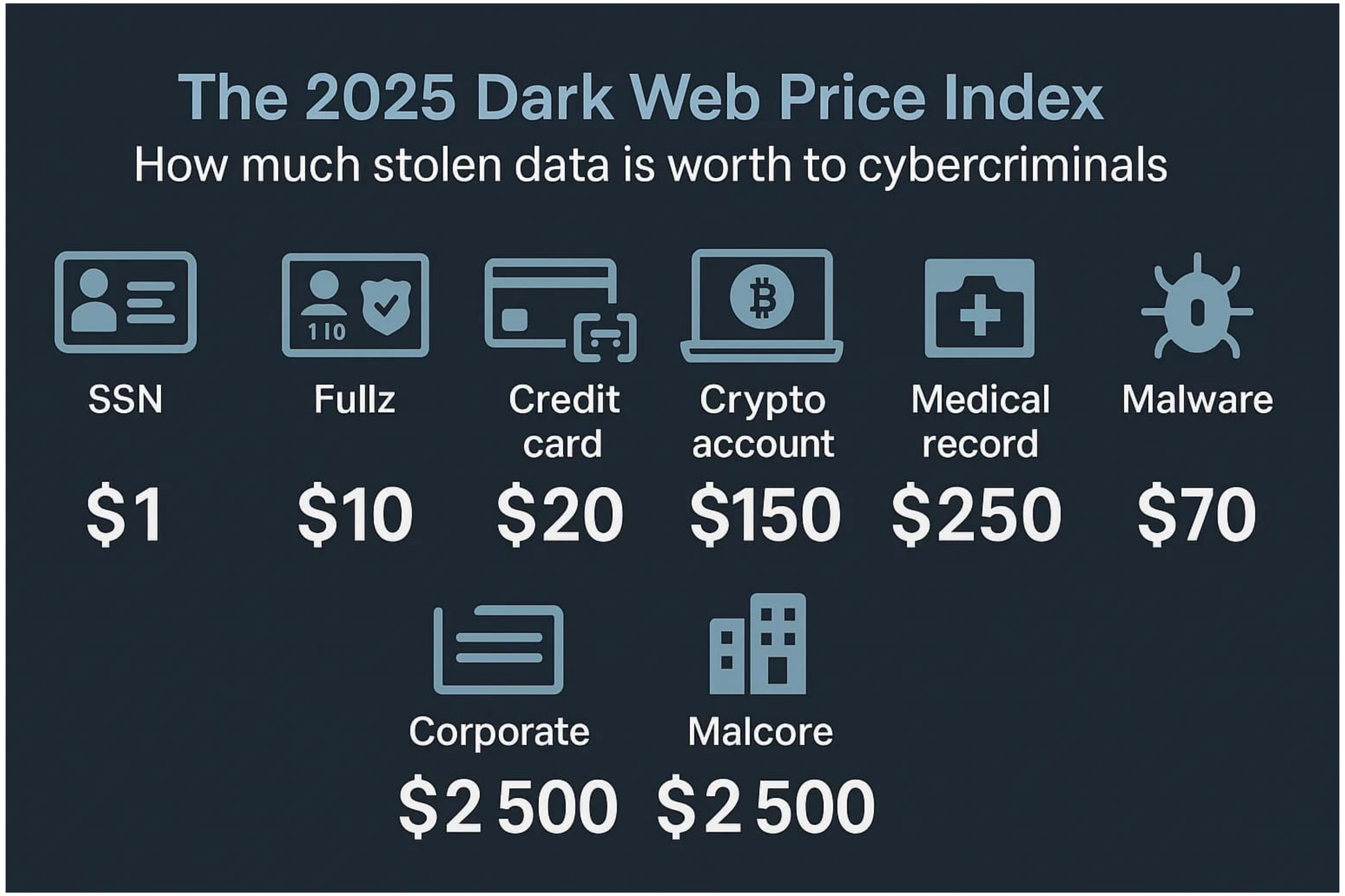


Don't want to give out your phone number?

No Problem. Use Ours.

Economics of Security

- Everything has a price!



The 2025 Dark Web Price Index

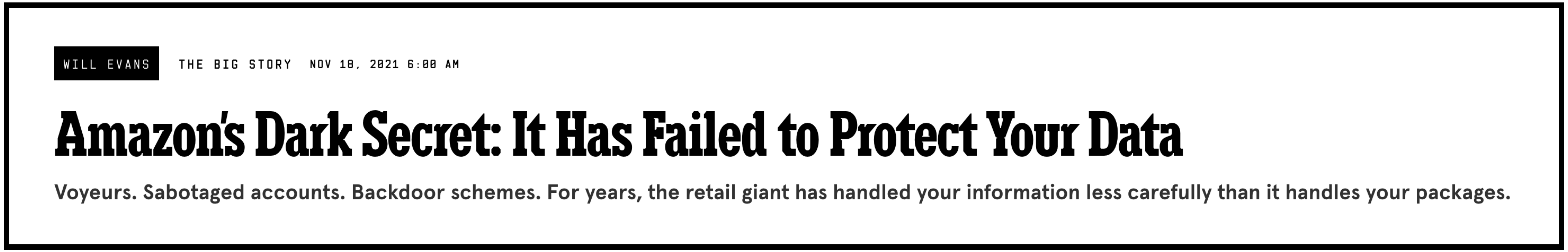
How much stolen data is worth to cybercriminals

SSN	Fullz	Credit card	Crypto account	Medical record	Malware
$1	$10	$20	$150	$250	$70

Corporate	Malcore
$2 500	$2 500

https://deepstrike.io/blog/dark-web-data-pricing-2025

Economics of Privacy

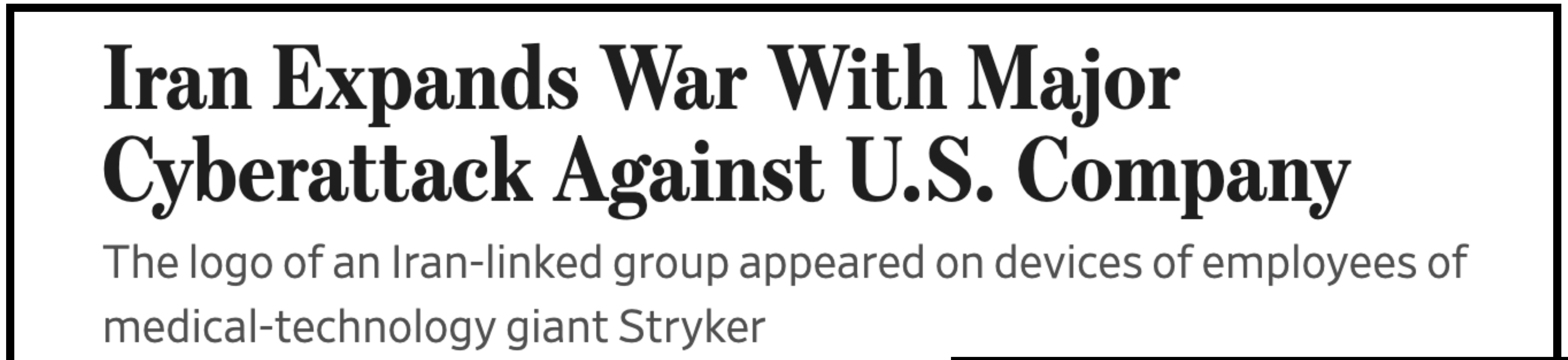


WILL EVANS THE BIG STORY NOV 18, 2021 6:00 AM

Amazon's Dark Secret: It Has Failed to Protect Your Data

Voyeurs. Sabotaged accounts. Backdoor schemes. For years, the retail giant has handled your information less carefully than it handles your packages.

Security and war are often linked

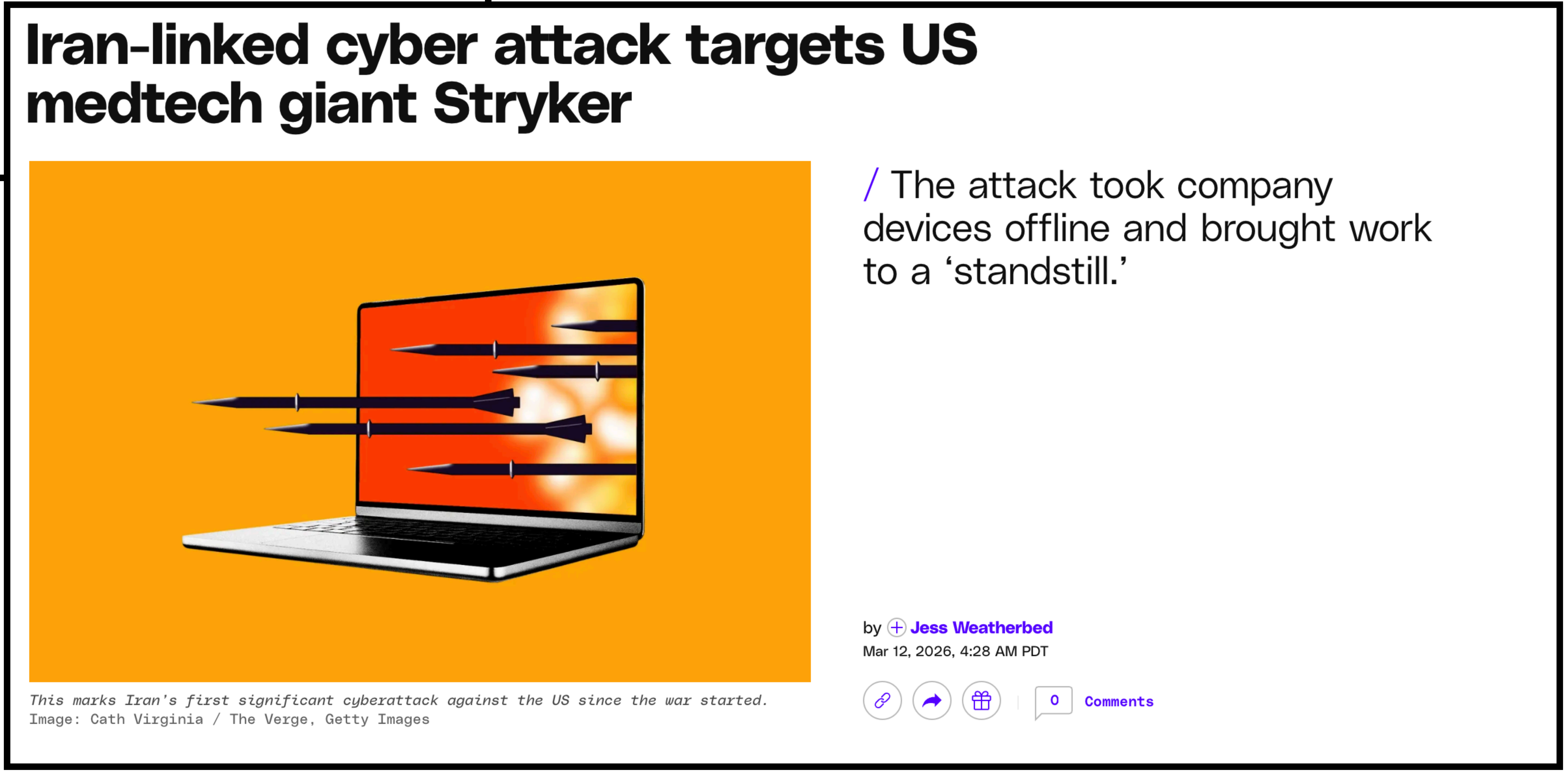


Iran Expands War With Major Cyberattack Against U.S. Company

The logo of an Iran-linked group appeared on devices of employees of medical-technology giant Stryker

By James Rundle and Dustin Volz

WSJ PRO Updated March 11, 2026 6:14 pm ET



Iran-linked cyber attack targets US medtech giant Stryker

/ The attack took company devices offline and brought work to a 'standstill.'

This marks Iran's first significant cyberattack against the US since the war started.
Image: Cath Virginia / The Verge, Getty Images

by Jess Weatherbed
Mar 12, 2026, 4:28 AM PDT

0 Comments

Practical security tips for your life

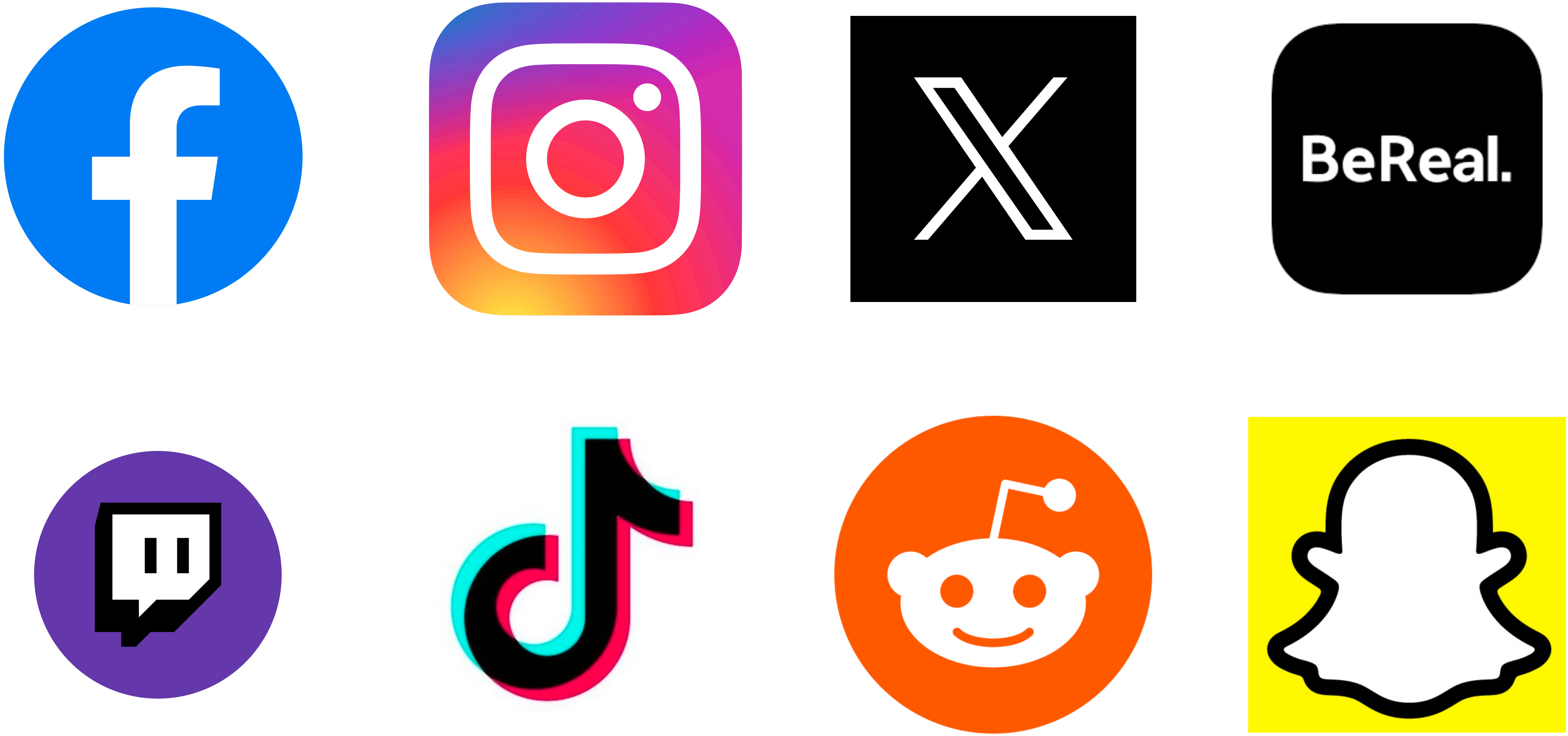
- Use a password manger!!!
  - Helps to limit password re-use, plus you will feel great when you never have to remember a password ever again
- Turn on multifactor authentication (**at least** for your e-mail and bank)
  - Authenticator apps are better than SMS if you have a choice of a 2nd factor
- Think about what you're really selling when you buy stuff from Internet companies… and whether you think it's worth it
- Invest in regular data backups (backup your machine **on a hard drive** at least once a day)
  - You never know when things are going to go south

SETs

Sociotechnical Cybersecurity

Let’s poll the room

Let’s poll the room

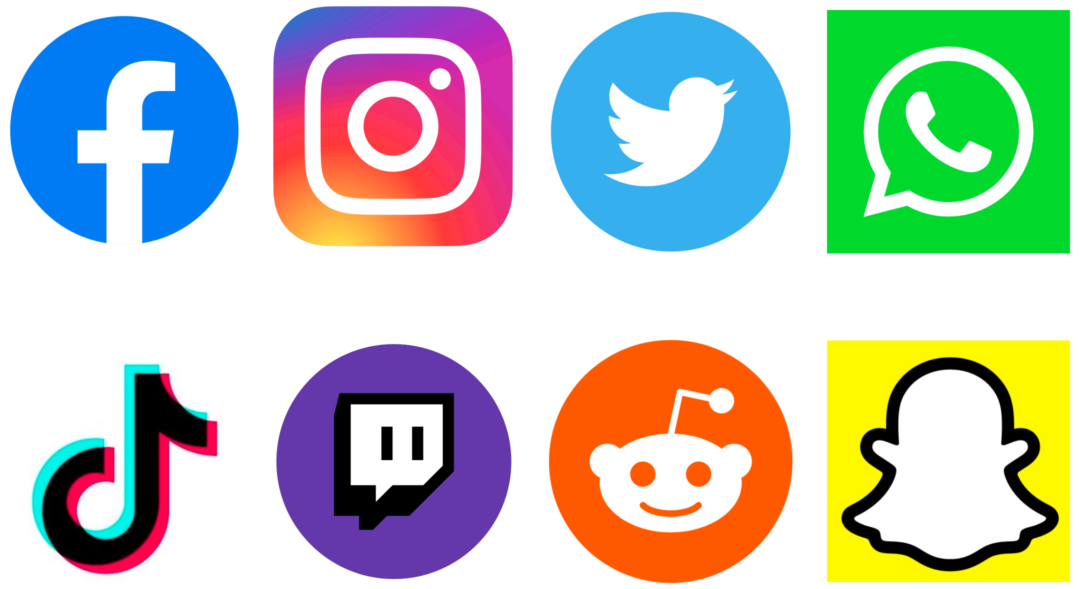


The rise of sociotechnical systems

Technology that interacts with personal, community, or societal interests

The rise of sociotechnical systems

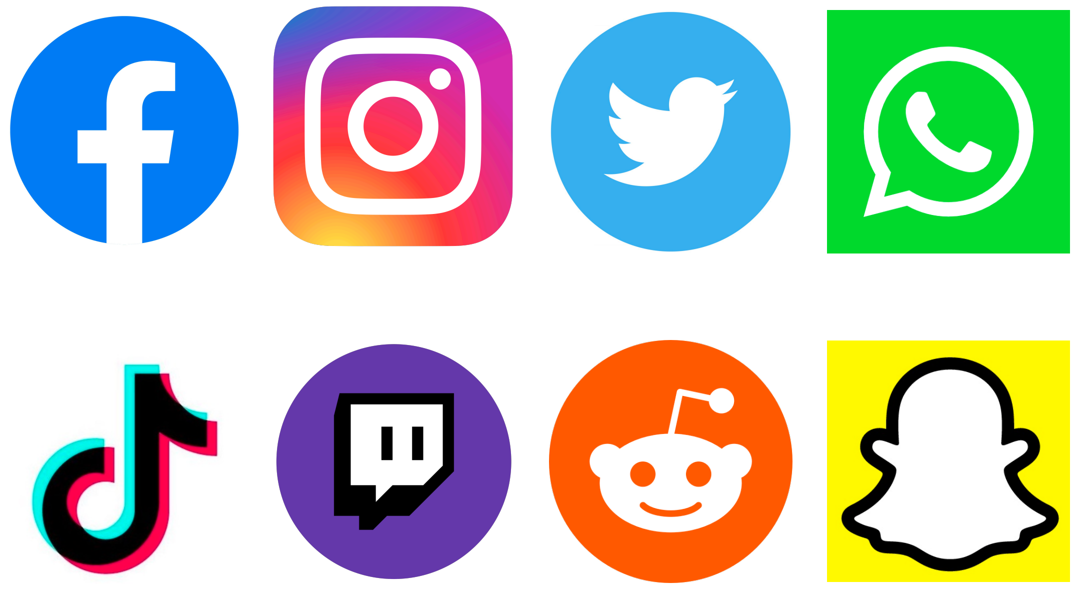
Technology that interacts with personal, community, or societal interests



Social Media

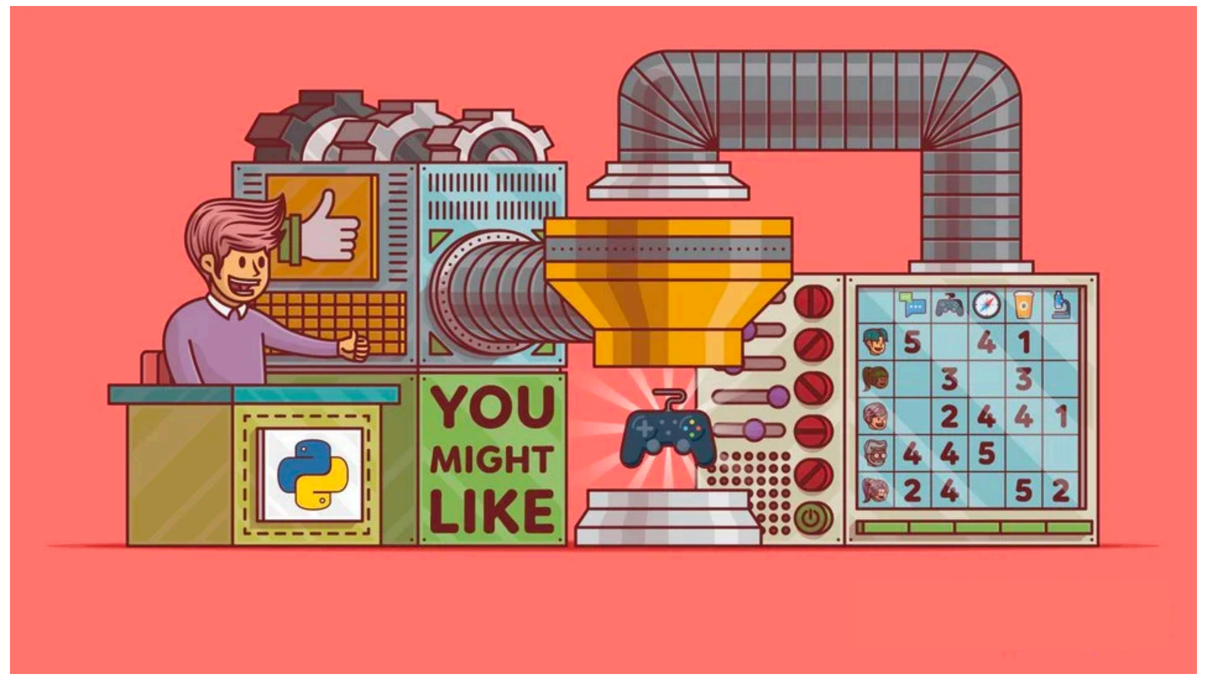
The rise of sociotechnical systems

Technology that interacts with personal, community, or societal interests



Social Media

Internet of Things



Recommender Systems

Sociotechnical Cybersecurity

How do computer systems fail in the presence of an adversary?

cal

Cybersecurity

Defining Sociotechnical Cybersecurity

The study of how an adversary can use a computer system to cause societal-level harms.

Sociotechnical systems enable security + safety threats

Security + safety challenges emerge at scale

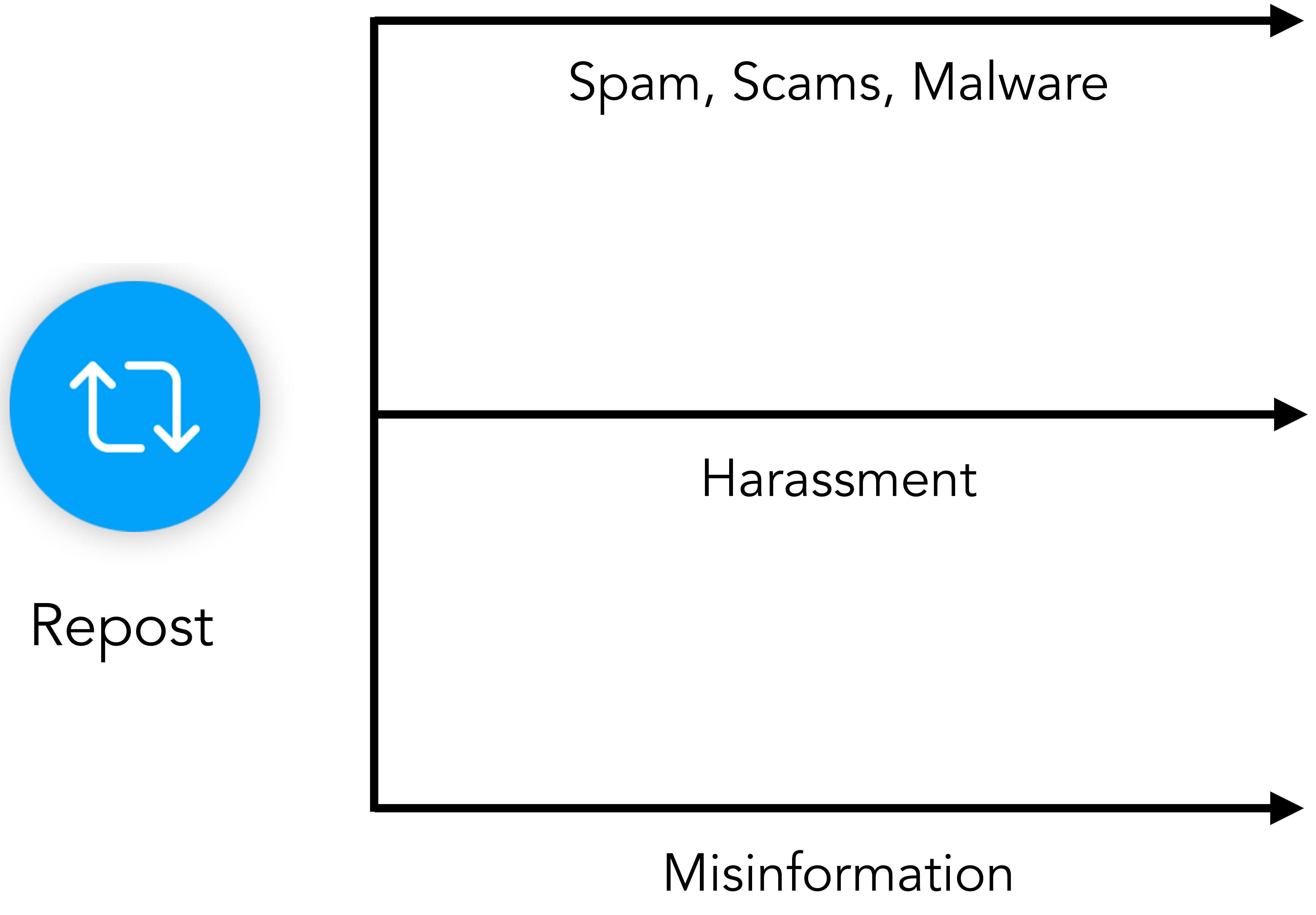
Sociotechnical systems enable security + safety threats

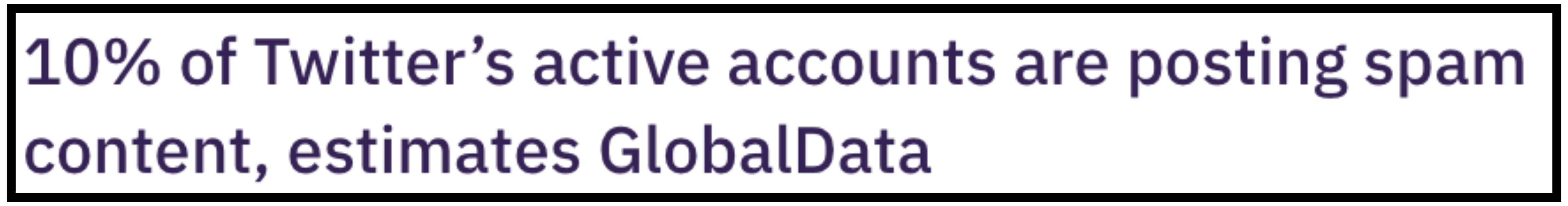
Security + safety challenges emerge at scale

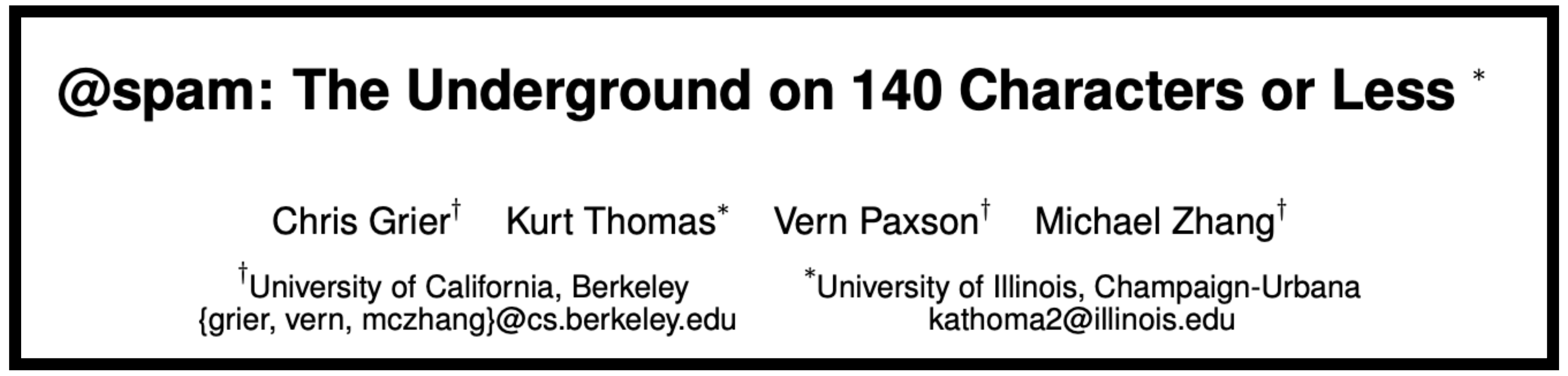
Repost

Sociotechnical systems enable security + safety threats

Security + safety challenges emerge at scale



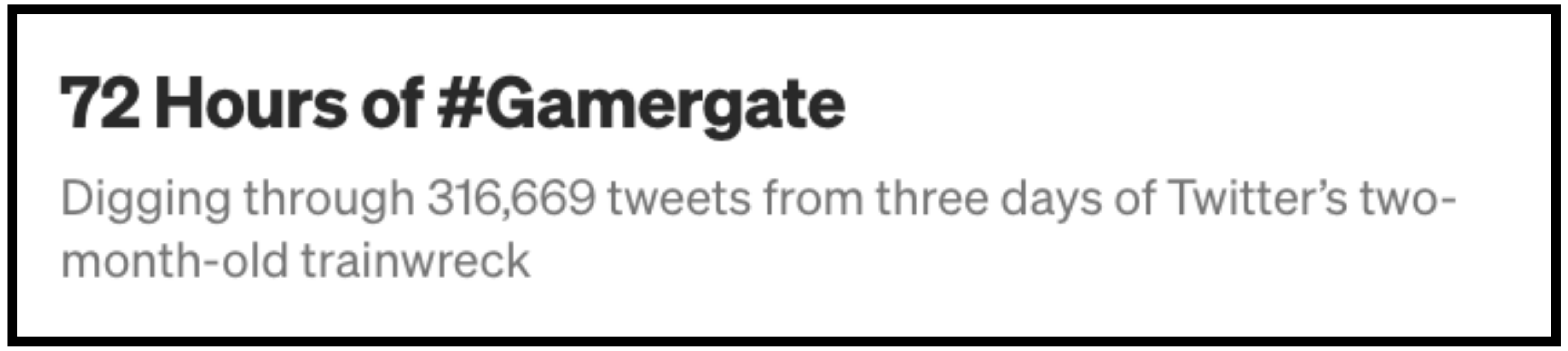

10% of Twitter’s active accounts are posting spam content, estimates GlobalData


@spam: The Underground on 140 Characters or Less [*]

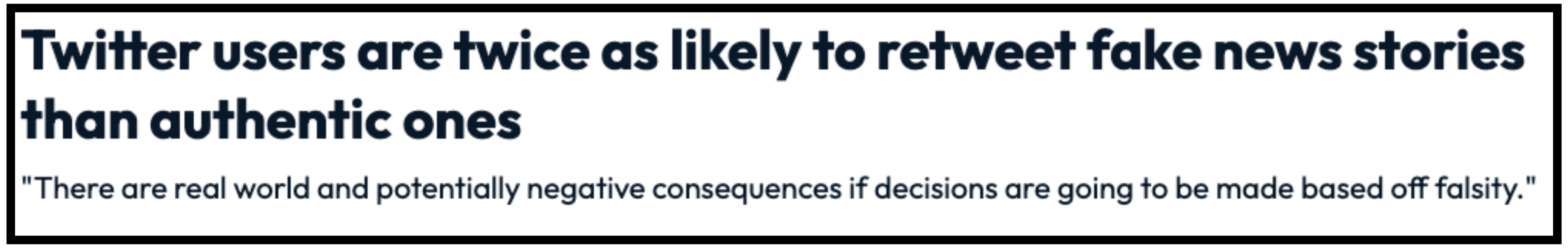
Chris Grier[†] Kurt Thomas[*] Vern Paxson[†] Michael Zhang[†]

[†]University of California, Berkeley
{grier, vern, mczhang}@cs.berkeley.edu

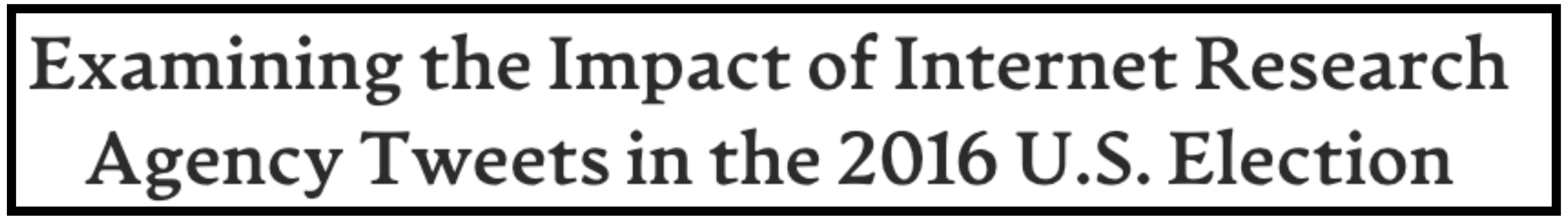
[*]University of Illinois, Champaign-Urbana
kathoma2@illinois.edu


72 Hours of #Gamergate

Digging through 316,669 tweets from three days of Twitter’s two-month-old trainwreck


Twitter users are twice as likely to retweet fake news stories than authentic ones

"There are real world and potentially negative consequences if decisions are going to be made based off falsity."


Examining the Impact of Internet Research Agency Tweets in the 2016 U.S. Election

So ats

Re

spam

stories

ed off falsity."

rch

n

Goal: Build safer and more secure sociotechnical systems

How do we reason about security and safety in such complex systems?

Security + safety through measurement

Deductively reasoning about systems is hard… so we need inductive approaches

Three primary techniques:

1. Large-scale measurements to study the scope, nature, and nuances of novel threats at scale

2. Human-centered studies (surveys, interviews) to understand how people experience security + safety threats

3. Design of systems, interventions, and defenses against these types of threats

Security + safety through measurement

IoT Security

Understanding the Mirai Botnet
(**USENIX '17**)

Skill Squatting Attacks on Amazon Alexa
(**USENIX '18**)

SoK: "Plug & Pray Today – Device Insecurity in USB Versions 1 through C
(**IEEE S&P '18**)

All Things Considered: An Analysis of IoT Devices on Home Networks
(**USENIX '19**)

Network Security

Security Challenges in an Increasingly Tangled Web
(**WWW '17**)

Tracking Certificate Misissuance in the Wild
(**IEEE S&P '18**)

Measuring Identity Confusion with URLs
(**CHI '20**)

Measuring DNS-over-HTTPS Performance Around the World
(**IMC '21**)

Detecting DNS Manipulation with TLS Certificate
(**PETS '23**)

Misinformation

On the Infrastructure Providers that Support Misinformation Websites
(**ICWSM '22**)

No Calm in the Storm: Investigate QAnon Website Relationships
(**ICWSM '22**)

Specious Sites: Tracking the Spread and Sway of Spurious News Stories at Scale
(**IEEE S&P '24**)

Happenstance: Utilizing Semantic Search to Track Russian State Media Narratives about the Russo-Ukrainian War on Reddit
(**ICWSM '23**)

Online Abuse

SoK: Hate, Harassment, and the Changing Landscape of Online Abuse
(**IEEE S&P '21**)

Designing Toxic Content Classification for a Diversity of Perspectives
(**SOUPS '21**)

Hate Raids on Twitch: Echoes of the Past, New Modalities, and Implications for Platform Governance
(**CSCW '23**) – **Best Paper**

Understanding the Behaviors of Toxic Accounts on Reddit
(**WWW '23**)

Specious Sites: Tracking the Spread and Sway of Spurious News Stories at Scale

joint w/ **Hans Hanley,** Zakir Durumeric

UC San Diego

Stanford University

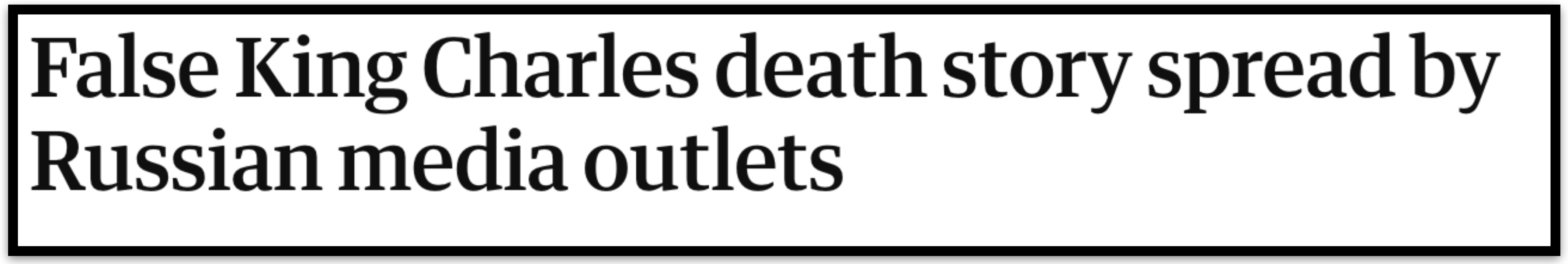
King Charles Is Not Dead

- In March 2024, several Russian news outlets began writing and spreading a rumor that King Charles III had died suddenly

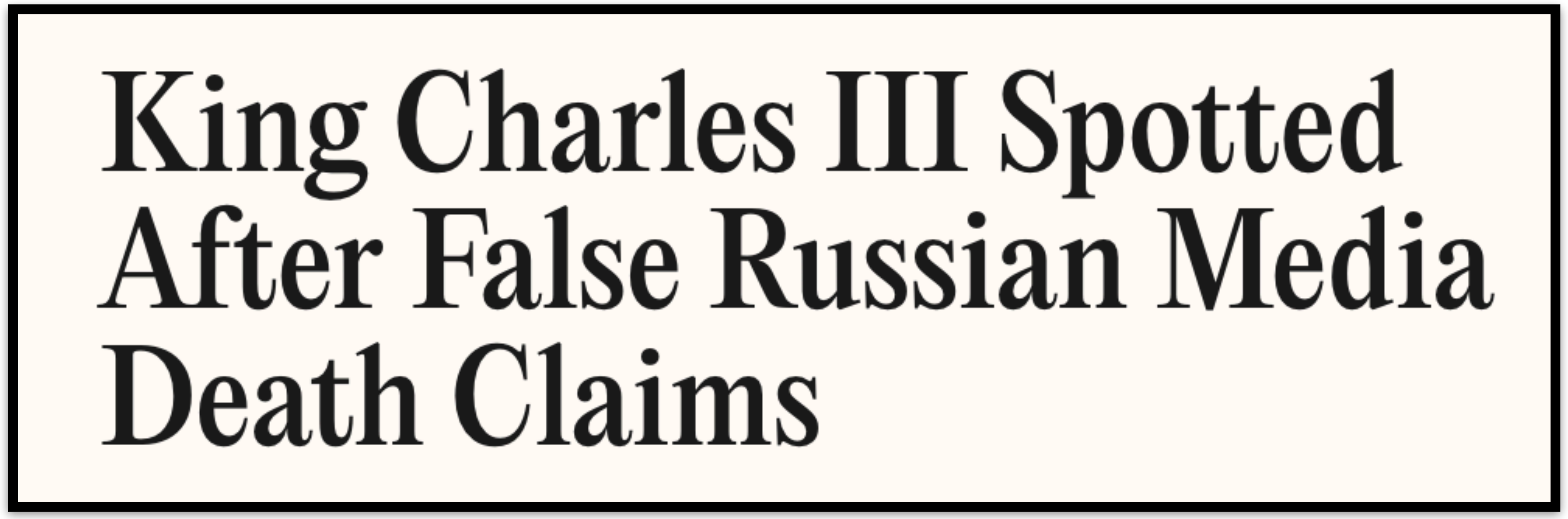
- Vedomosti, Open Ukraine, Uncle Slava, Sputniknews, and Readovka all spread the rumor through their websites

- Quickly picked up by high volume Telegram channels, started spreading online

- Event prompted Buckingham Palace to have to respond noting that the King is alive, well, and good…



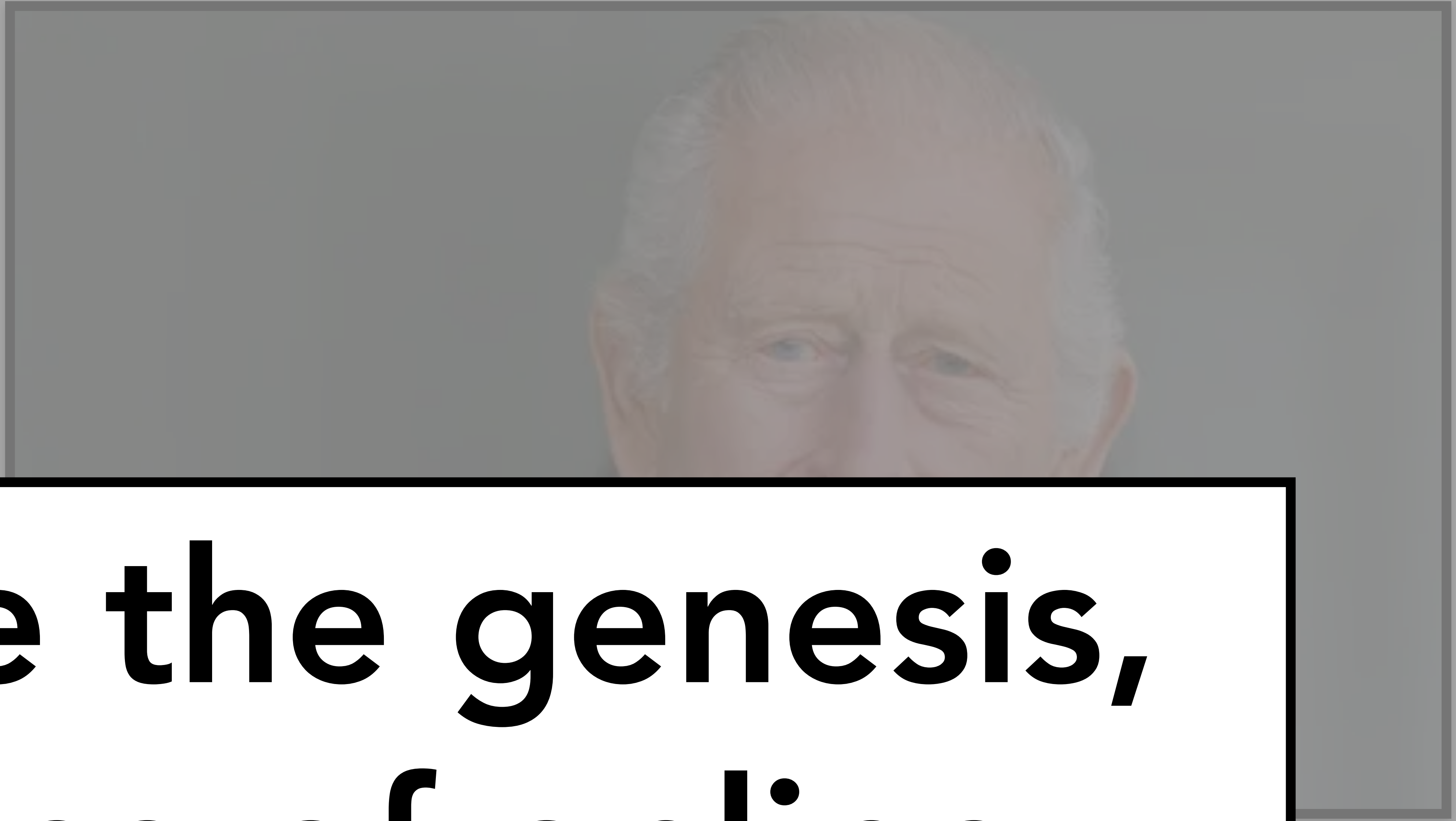
False King Charles death story spread by Russian media outlets



King Charles III Spotted After False Russian Media Death Claims

King Charles Is Not Dead

- In March 2024, several Russian news outlets began writing and spreading a rumor that



How can we measure the genesis, spread, and influence of online misinformation narratives at scale?

d by

channels, started spreading online

- Event prompted Buckingham Palace to have to respond noting that the King is alive, well, and good…

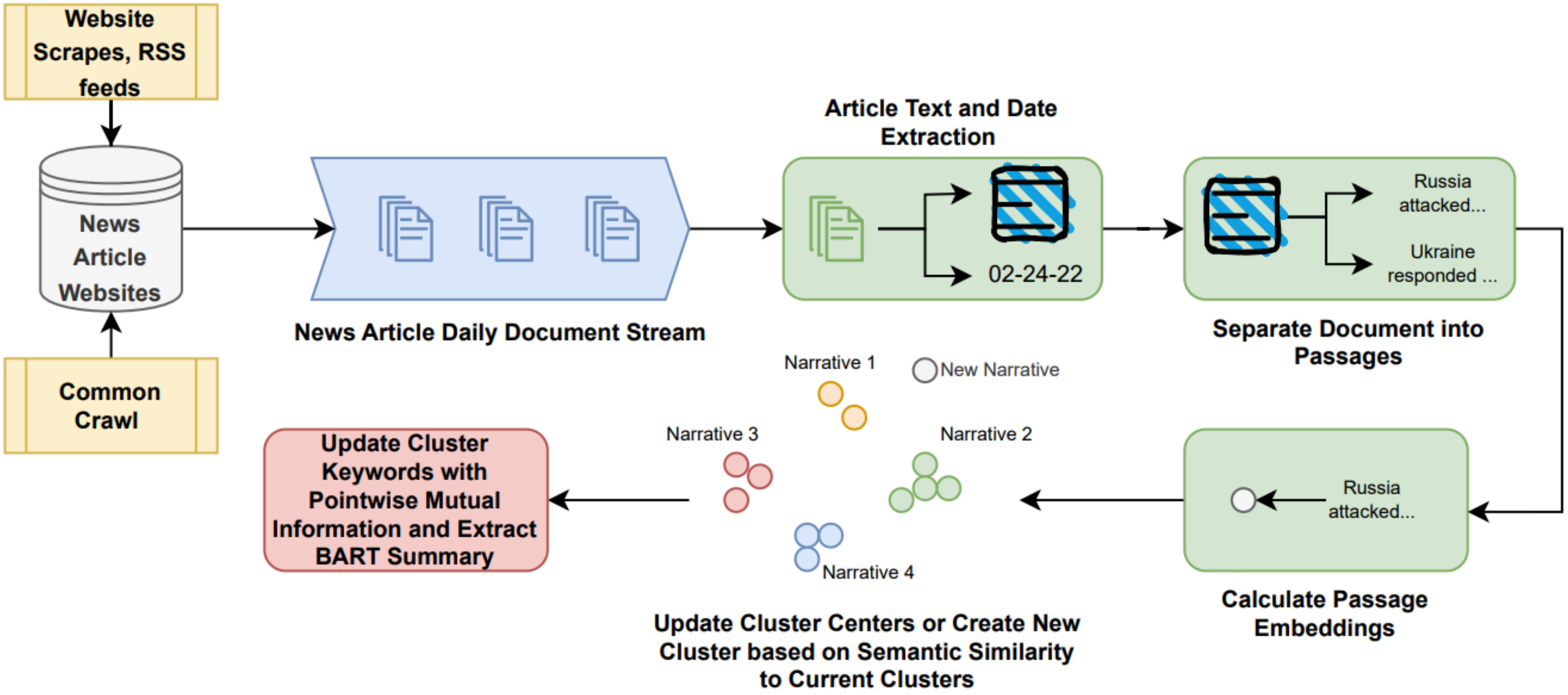
King Charles III Spotted After False Russian Media Death Claims

First, what’s a narrative?

- “Collections of information that seek to address the same *event* or *issue*.”
    - “Electron fraud in the 2020 U.S. election”
    - “COVID-19 vaccine leading to mass death”
- Not all related topics are similar, e.g.,
    - “US funds Ukrainian War” **not the same narrative** as “Russia attacks Ukraine”

Specious Sites

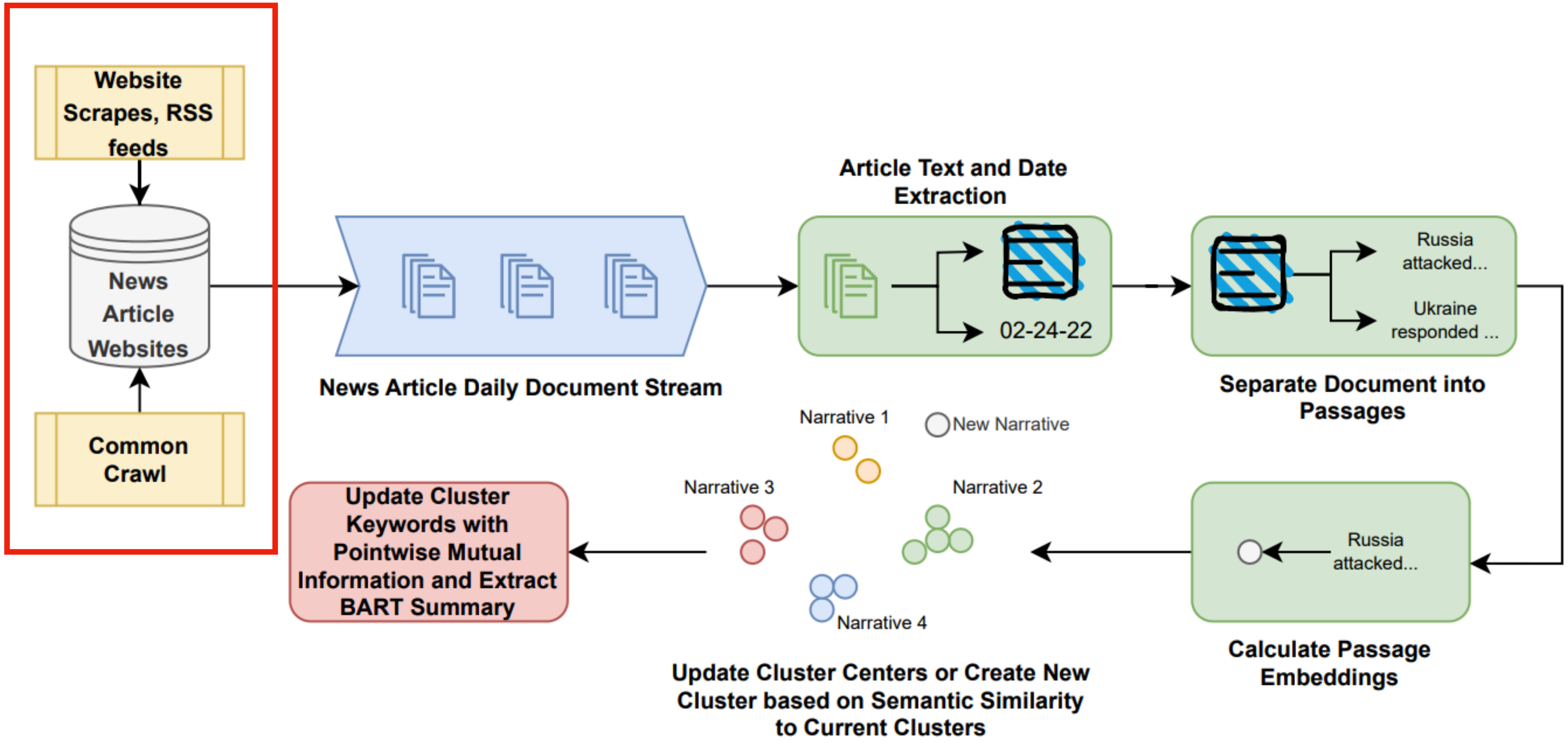
Goal of the work: Build automated and programmatic approach for tracking news narratives on the web



Website Scrapes, RSS feeds
News Article Websites
Common Crawl
News Article Daily Document Stream
Article Text and Date Extraction
02-24-22
Russia attacked...
Ukraine responded ...
Separate Document into Passages
Russia attacked...
Calculate Passage Embeddings
Narrative 1
New Narrative
Narrative 3
Narrative 2
Narrative 4
Update Cluster Centers or Create New Cluster based on Semantic Similarity to Current Clusters
Update Cluster Keywords with Pointwise Mutual Information and Extract BART Summary

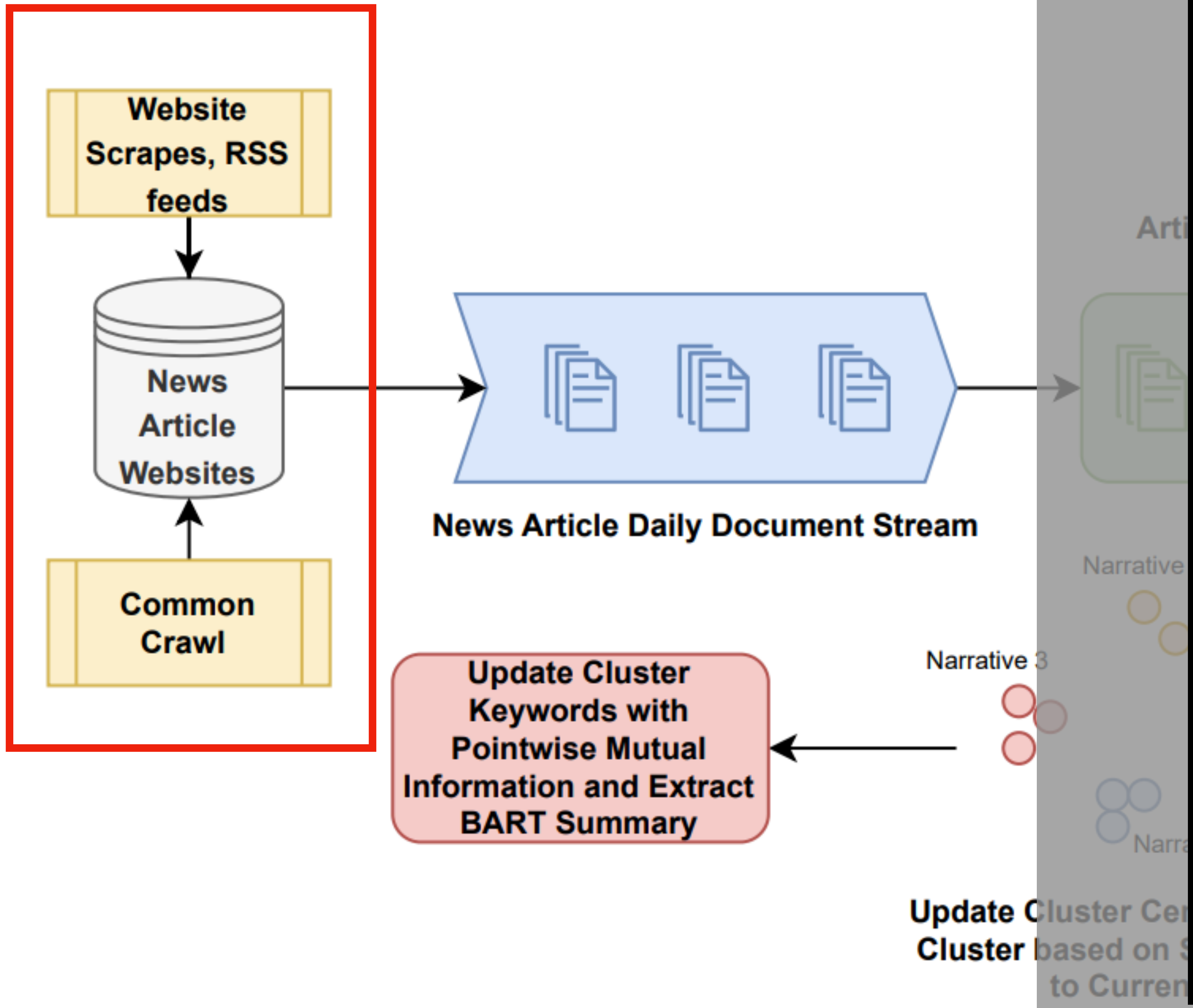
Specious Sites

Goal of the work: Build automated and programmatic approach for tracking news narratives on the web


Website Scrapes, RSS feeds
News Article Websites
Common Crawl
News Article Daily Document Stream
Article Text and Date Extraction
02-24-22
Russia attacked...
Ukraine responded ...
Separate Document into Passages
Russia attacked...
Calculate Passage Embeddings
Narrative 1
New Narrative
Narrative 3
Narrative 2
Narrative 4
Update Cluster Centers or Create New Cluster based on Semantic Similarity to Current Clusters
Update Cluster Keywords with Pointwise Mutual Information and Extract BART Summary

Specious Sites

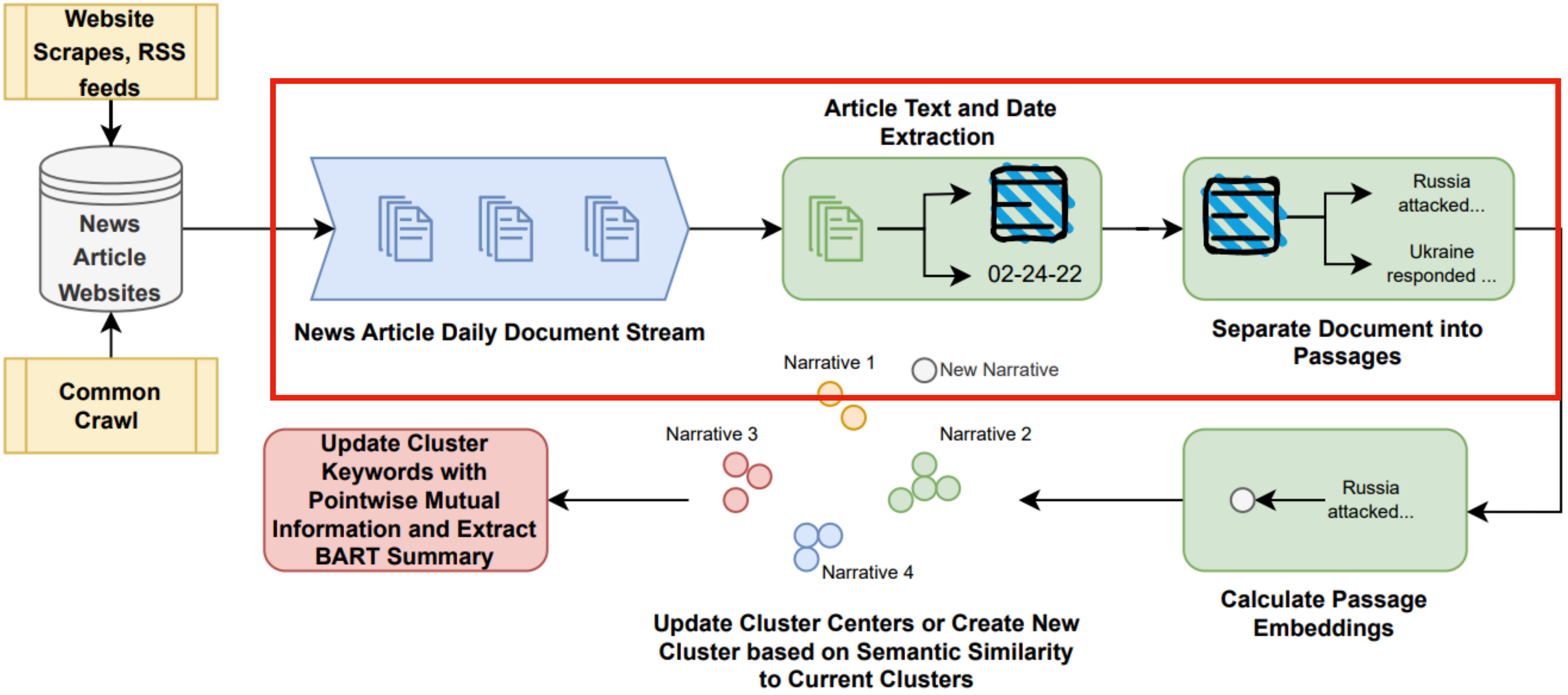
Goal of the work: Build automated and p
news narratives on the web


Website Scrapes, RSS feeds
News Article Websites
Common Crawl
News Article Daily Document Stream
Arti
Narrative
Narrative 3
Narra
Update Cluster Keywords with Pointwise Mutual Information and Extract BART Summary
Update Cluster Cen
Cluster based on
to Curren

- Conducted daily scrapes of **1334** unreliable and politically biased news websites
- Includes “politically biased,” “misinformation,” “disinformation,” “conspiracy,” “fake news,” or “state-based propaganda” as labeled by previous studies
- Collected **1.92M** articles from **January 1, 2022 — November 1, 2022**

Specious Sites

Goal of the work: Build automated and programmatic approach for tracking news narratives on the web

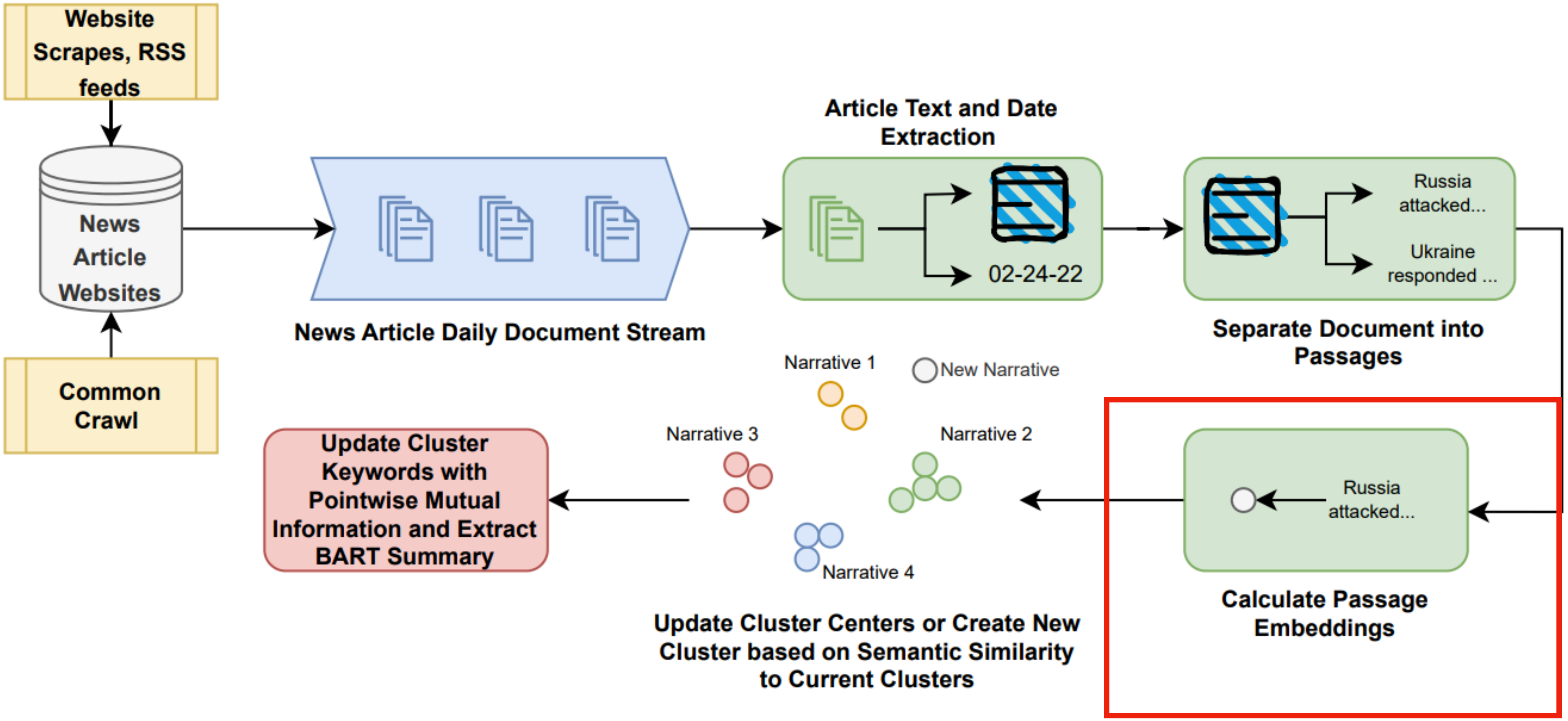

Website Scrapes, RSS feeds
News Article Websites
Common Crawl
Article Text and Date Extraction
02-24-22
News Article Daily Document Stream
Russia attacked...
Ukraine responded ...
Separate Document into Passages
Narrative 1
New Narrative
Narrative 3
Narrative 2
Narrative 4
Update Cluster Keywords with Pointwise Mutual Information and Extract BART Summary
Russia attacked...
Calculate Passage Embeddings
Update Cluster Centers or Create New Cluster based on Semantic Similarity to Current Clusters

Segmenting and Preprocessing News Articles

- Next, tons of preprocessing
  - Endless libraries, custom website parsers, etc., to identify *article text* and *metadata* (e.g., date, author, etc.)
- **Split articles into ~100-word *passages*.**
  - Articles can (and often do) reference multiple *events*, we want to capture that granularity
  - We needed to balance granularity with performance (e.g., sentences vs. documents)
  - SOTA embedding models (at the time) had a limited context window so we went with what worked

Specious Sites

Goal of the work: Build automated and programmatic approach for tracking news narratives on the web


Website Scrapes, RSS feeds
News Article Websites
Common Crawl
News Article Daily Document Stream
Article Text and Date Extraction
02-24-22
Russia attacked...
Ukraine responded ...
Separate Document into Passages
Russia attacked...
Calculate Passage Embeddings
Narrative 1
New Narrative
Narrative 2
Narrative 3
Narrative 4
Update Cluster Centers or Create New Cluster based on Semantic Similarity to Current Clusters
Update Cluster Keywords with Pointwise Mutual Information and Extract BART Summary

Embedding Passages

- We embedded each of our resultant **25M passages** with fine-tuned version of MPNet, a semantic similarity model from Microsoft

Ukrainian Neo-Nazis

Example passage 1: A subversive group of militants of the Ukrainian Neo-nazi Azov formation attacked Russian troops.

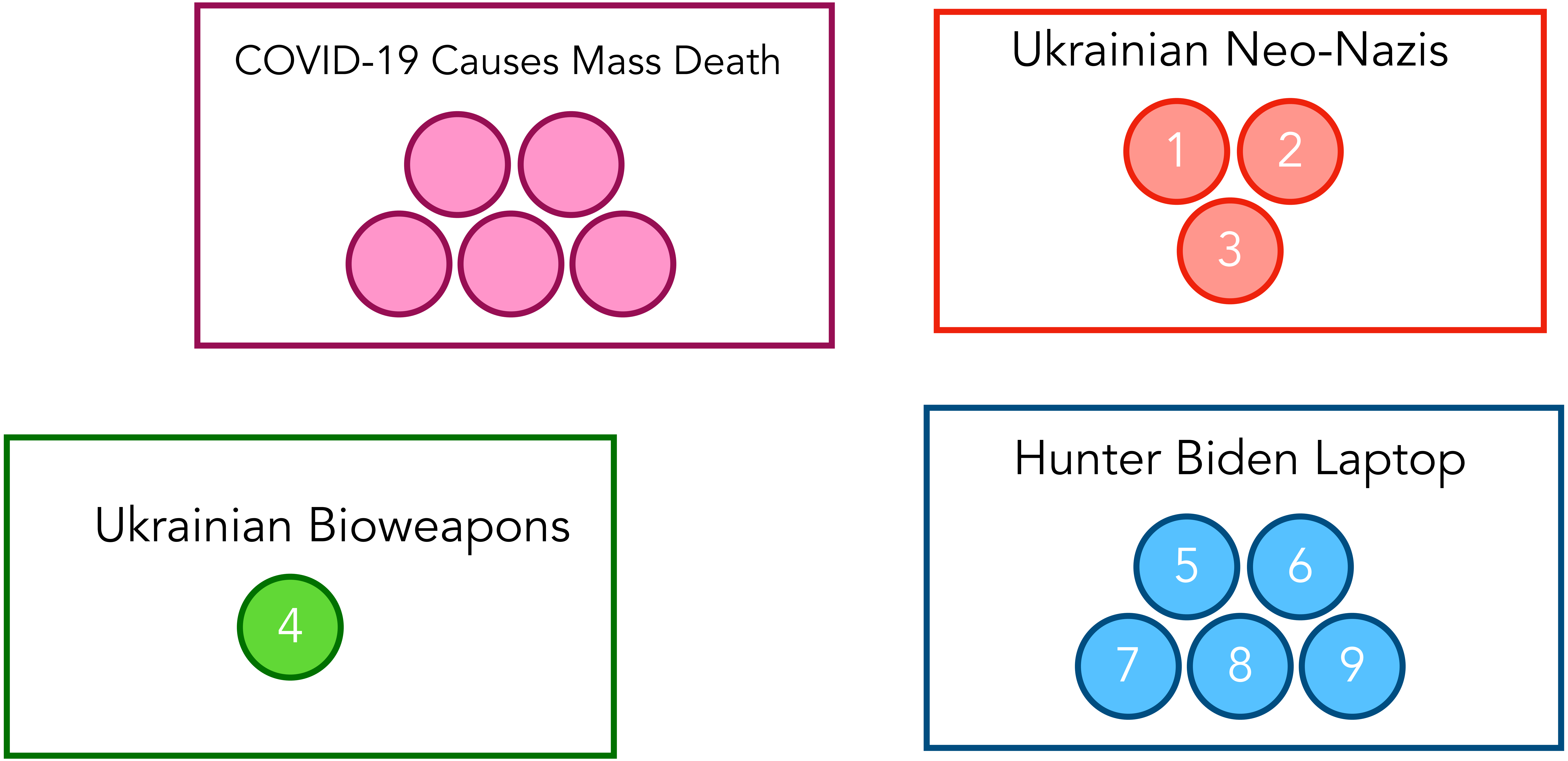
Example passage 2: Asov is a far-right organization that welcomes all sorts of neonazis.

Example passage 3: Ukraine neo-nazi battalion has built a state within a state.

Ukrainian Bioweapons

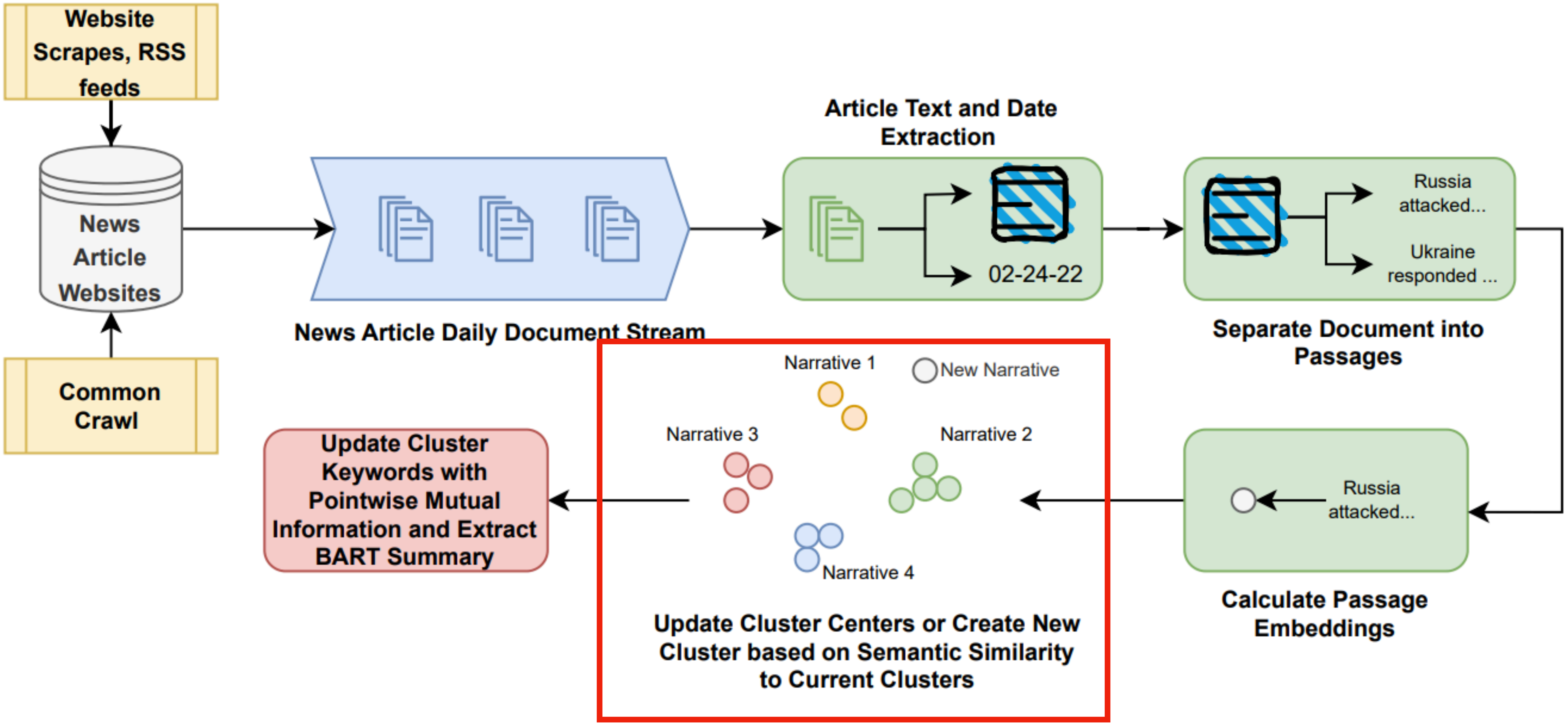
Example passage 4: Ukraine was developing biological weapons with the assistance of the US government.

Semantic Embeddings



Specious Sites

Goal of the work: Build automated and programmatic approach for tracking news narratives on the web



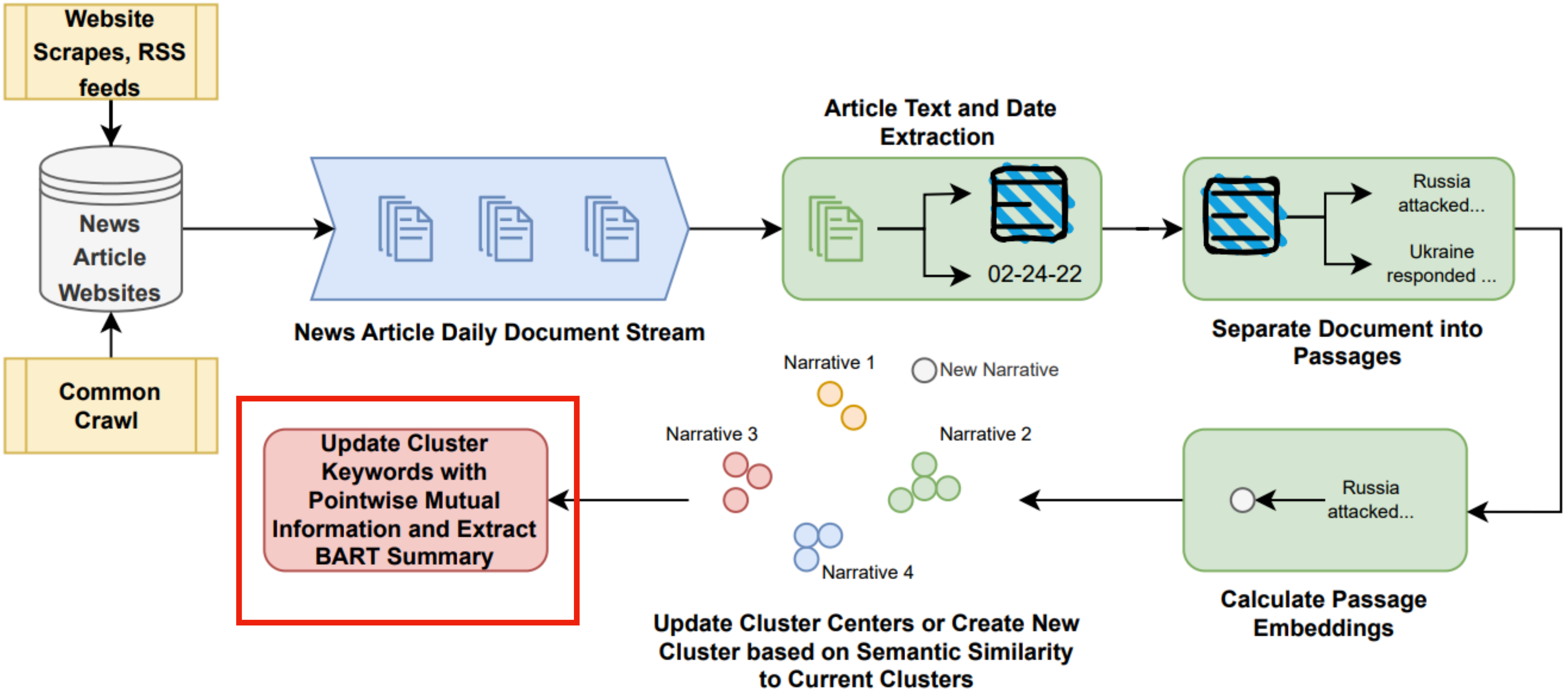
Website Scrapes, RSS feeds
News Article Websites
Common Crawl
News Article Daily Document Stream
Article Text and Date Extraction
02-24-22
Russia attacked...
Ukraine responded ...
Separate Document into Passages
Russia attacked...
Calculate Passage Embeddings
Narrative 1
New Narrative
Narrative 3
Narrative 2
Narrative 4
Update Cluster Centers or Create New Cluster based on Semantic Similarity to Current Clusters
Update Cluster Keywords with Pointwise Mutual Information and Extract BART Summary

Clustering Similar Passages

- Clustered similar passages together via DP-Means
  - Non-parametric K-means; no specific # of clusters a priori
  - Group clusters together based on cosine similarity
- Started with initial set of clusters, then updated clusters based on new embeddings found in each new day of the dataset
  - Simulating a "real-time" clustering
- Ultimately ended up with **52K** clusters of narratives among set of unreliable news websites — took ~**1.5 days using an A6000 GPU**

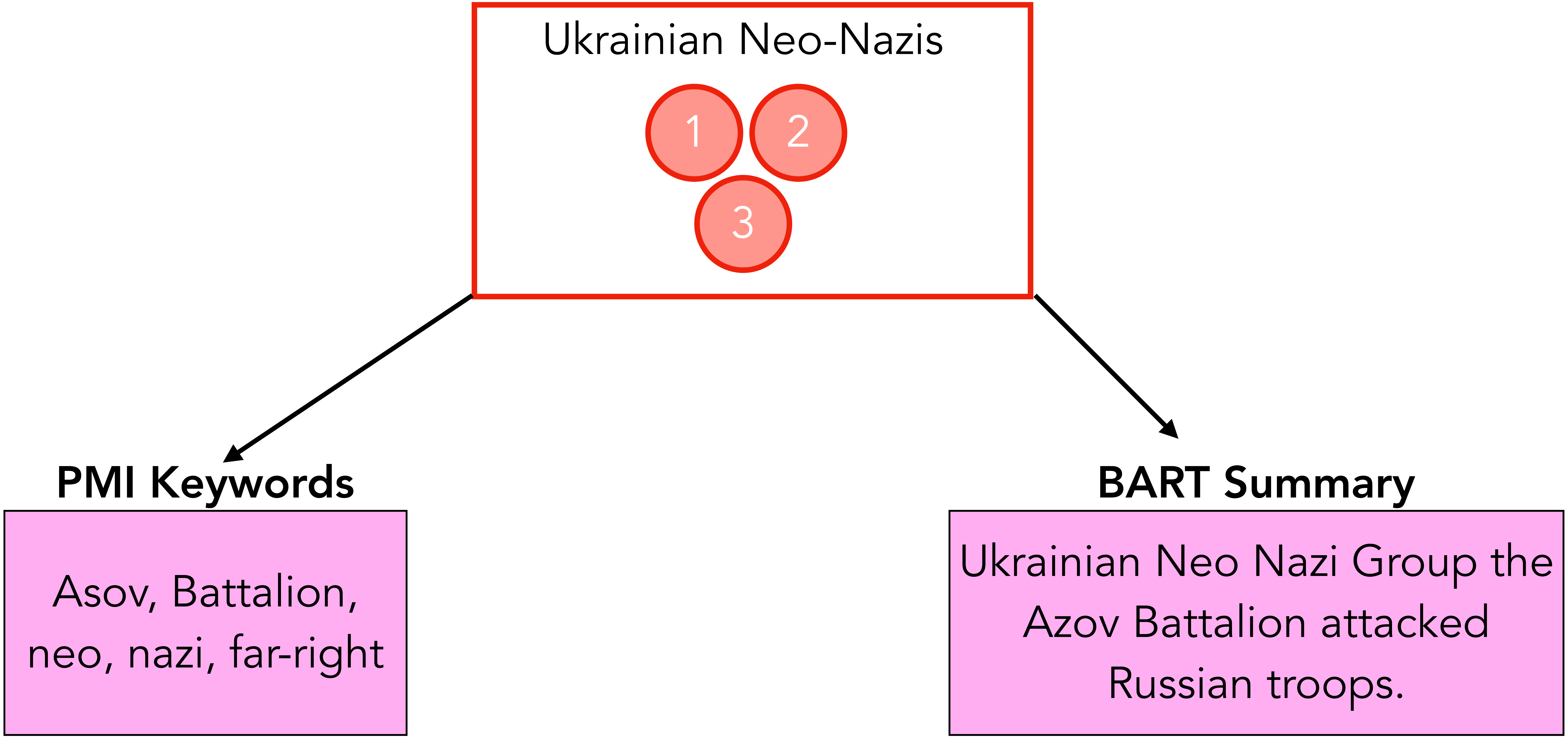
Specious Sites

Goal of the work: Build automated and programmatic approach for tracking news narratives on the web



Website Scrapes, RSS feeds
News Article Websites
Common Crawl
News Article Daily Document Stream
Article Text and Date Extraction
02-24-22
Russia attacked...
Ukraine responded ...
Separate Document into Passages
Russia attacked...
Calculate Passage Embeddings
Narrative 1
New Narrative
Narrative 3
Narrative 2
Narrative 4
Update Cluster Centers or Create New Cluster based on Semantic Similarity to Current Clusters
Update Cluster Keywords with Pointwise Mutual Information and Extract BART Summary

Summarizing Cluster Details



Ukrainian Neo-Nazis

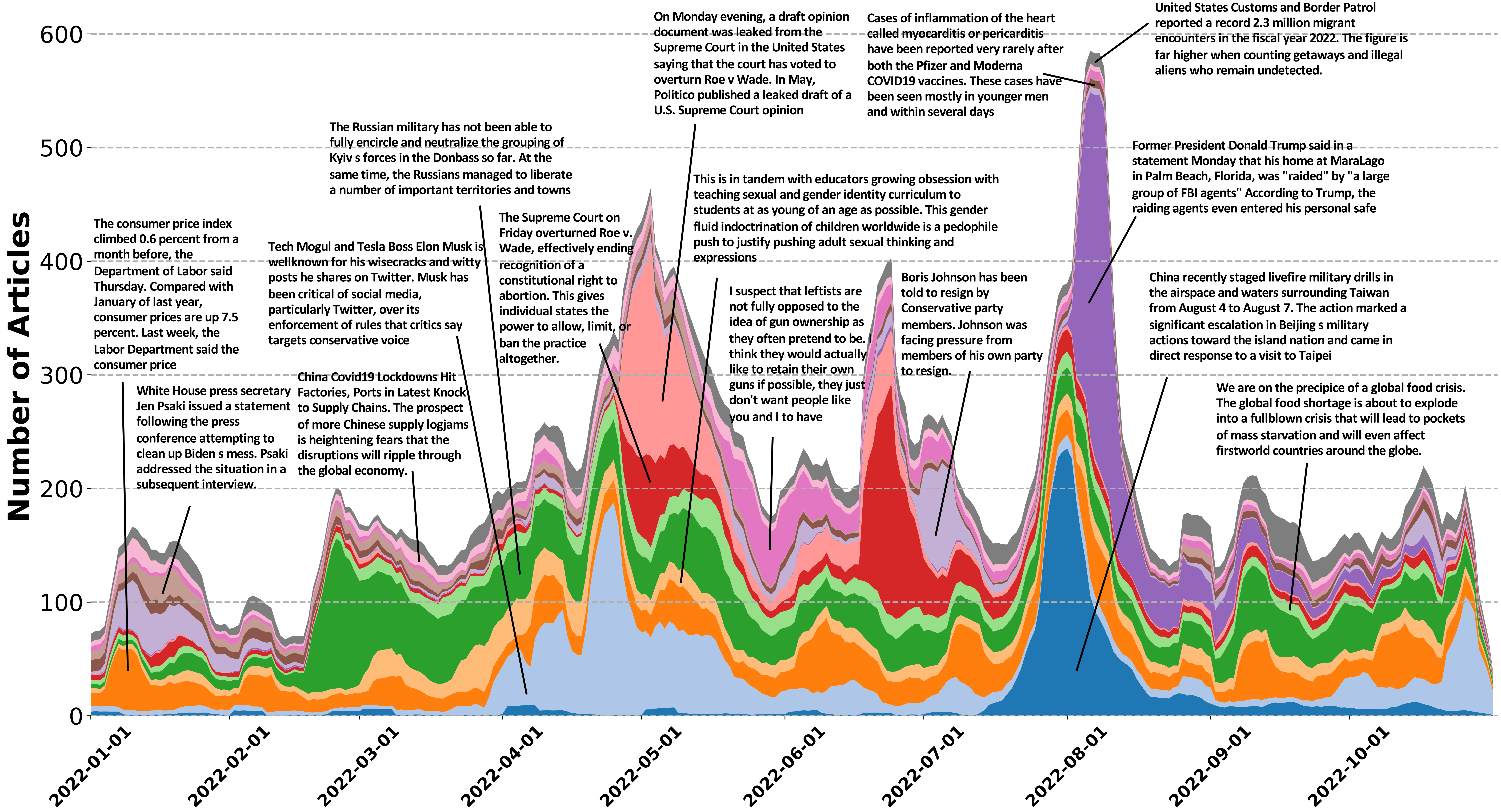
PMI Keywords

Asov, Battalion, neo, nazi, far-right

BART Summary

Ukrainian Neo Nazi Group the Azov Battalion attacked Russian troops.

Top narratives in our study



The consumer price index climbed 0.6 percent from a month before, the Department of Labor said Thursday. Compared with January of last year, consumer prices are up 7.5 percent. Last week, the Labor Department said the consumer price

White House press secretary Jen Psaki issued a statement following the press conference attempting to clean up Biden s mess. Psaki addressed the situation in a subsequent interview.

Tech Mogul and Tesla Boss Elon Musk is wellknown for his wisecracks and witty posts he shares on Twitter. Musk has been critical of social media, particularly Twitter, over its enforcement of rules that critics say targets conservative voice

China Covid19 Lockdowns Hit Factories, Ports in Latest Knock to Supply Chains. The prospect of more Chinese supply logjams is heightening fears that the disruptions will ripple through the global economy.

The Russian military has not been able to fully encircle and neutralize the grouping of Kyiv s forces in the Donbass so far. At the same time, the Russians managed to liberate a number of important territories and towns

The Supreme Court on Friday overturned Roe v. Wade, effectively ending recognition of a constitutional right to abortion. This gives individual states the power to allow, limit, or ban the practice altogether.

On Monday evening, a draft opinion document was leaked from the Supreme Court in the United States saying that the court has voted to overturn Roe v Wade. In May, Politico published a leaked draft of a U.S. Supreme Court opinion

This is in tandem with educators growing obsession with teaching sexual and gender identity curriculum to students at as young of an age as possible. This gender fluid indoctrination of children worldwide is a pedophile push to justify pushing adult sexual thinking and expressions

I suspect that leftists are not fully opposed to the idea of gun ownership as they often pretend to be. I think they would actually like to retain their own guns if possible, they just don't want people like you and I to have

Cases of inflammation of the heart called myocarditis or pericarditis have been reported very rarely after both the Pfizer and Moderna COVID19 vaccines. These cases have been seen mostly in younger men and within several days

Boris Johnson has been told to resign by Conservative party members. Johnson was facing pressure from members of his own party to resign.

United States Customs and Border Patrol reported a record 2.3 million migrant encounters in the fiscal year 2022. The figure is far higher when counting getaways and illegal aliens who remain undetected.

Former President Donald Trump said in a statement Monday that his home at MaraLago in Palm Beach, Florida, was "raided" by "a large group of FBI agents" According to Trump, the raiding agents even entered his personal safe

China recently staged livefire military drills in the airspace and waters surrounding Taiwan from August 4 to August 7. The action marked a significant escalation in Beijing s military actions toward the island nation and came in direct response to a visit to Taipei

We are on the precipice of a global food crisis. The global food shortage is about to explode into a fullblown crisis that will lead to pockets of mass starvation and will even affect firstworld countries around the globe.

Estimating Narrative Origination + Amplification

- With this measurement, **we can identify where narratives originate** and **how influential the originator is in spreading new narratives**

- **Key idea:** Compare distributions of external articles when a given domain originates (posts a narrative first) to when it posts a story later in the narrative's lifecycle

  - …A *lot* more detail in the paper

Estimating Narrative Origination + Amplification

Domain	CrUX Rank	Weighted Avg. Delta	Effect Size
therightscoop.com	100K – 500K	0.684	1.758
weaselzippers.us	100 – 500K	0.674	1.519
toddstarnes.com	100 – 500K	0.576	1.475
nationalfile.com	500K – 1M	0.406	1.306
gellerreport.com	500K – 1M	0.359	1.259
usaanews.com	500K – 1M	0.3	1.238
infostormer.com	**500K – 1M**	**0.763**	**2.514**

Low popularity, fringe domains are best at originating narratives that are then picked up by more popular downstream outlets

Application: Tracking new narratives

- By monitoring day-over-day changes in clusters + narratives, we can quickly surface discussion and potential direction of narratives in near real-time

Paul Pelosi conspiracies surfaced by our system in last week of our study

Application: Fact-checking

- We compared our narrative tracking to articles fact-checked by Politifact, Reuters, and APNews
- Key question: **Can automated narrative tracking aid or augment existing fact-checking efforts?**

Application: Fact-checking

- Narratives can spread online for 49 – 83 days before fact-checked
  - Orgs fact-check after a story has "peaked" online; **check comes too late**
- Narrative tracking can surface stories to fact-checkers long before they peak, acting as a *proactive defense*

	Fact-checked stories	**Median Days to Fact Check**	**Median Days from Peak**
Politifact	6231	55	4
Reuters	9604	49	0
AP News	230	83	3

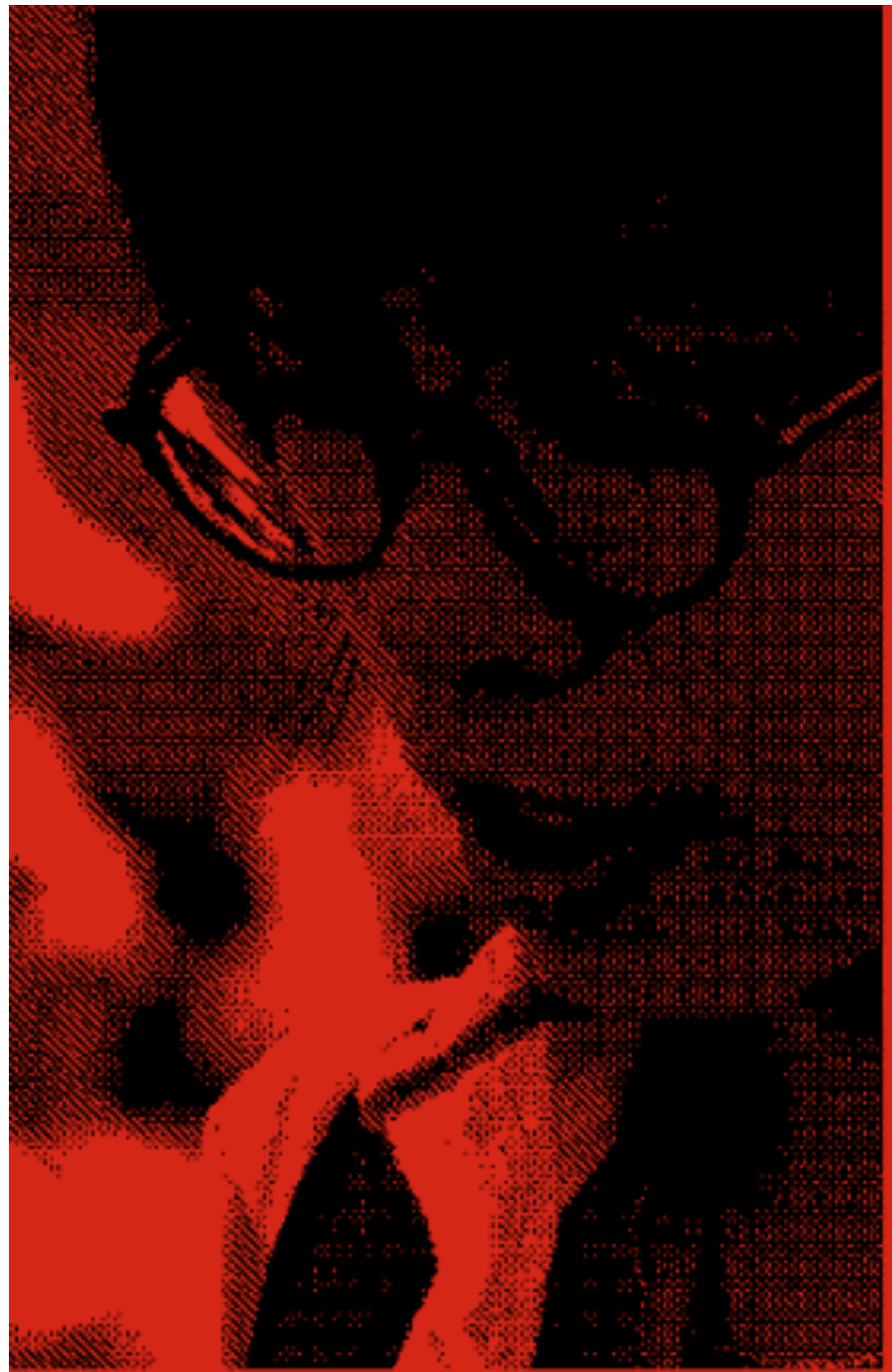
Future Plans

- Narrative tracking ***is possible*** with 2023-era tools, with some caveats
  - Clusters were decent, but sometimes imprecise (e.g., Monkeypox + COVID-19 outbreaks in NYC were frequently grouped together)
- Future work
  - Real-time monitoring (!!), domain discovery (!!), using **new AI tools;** can agents help with mis/disinformation discovery?

PressProtect: Helping Journalists Navigate Social Media in the Face of Online Harassment

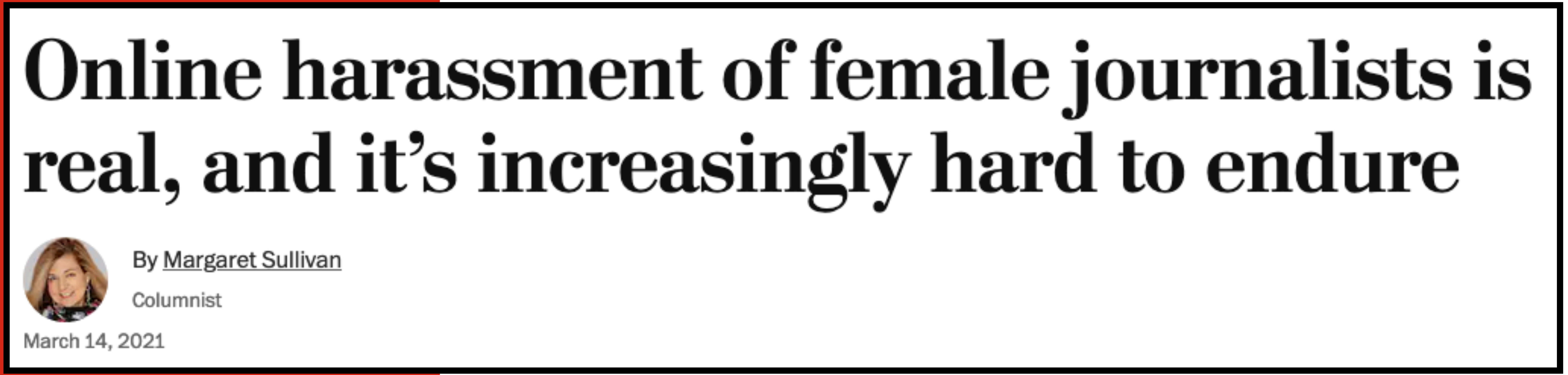
Catherine Han, Anne Li, Deepak Kumar, Zakir Durumeric

ACM CSCW 2024



Attacks and Harassment

The Impact on Female Journalists and Their Reporting

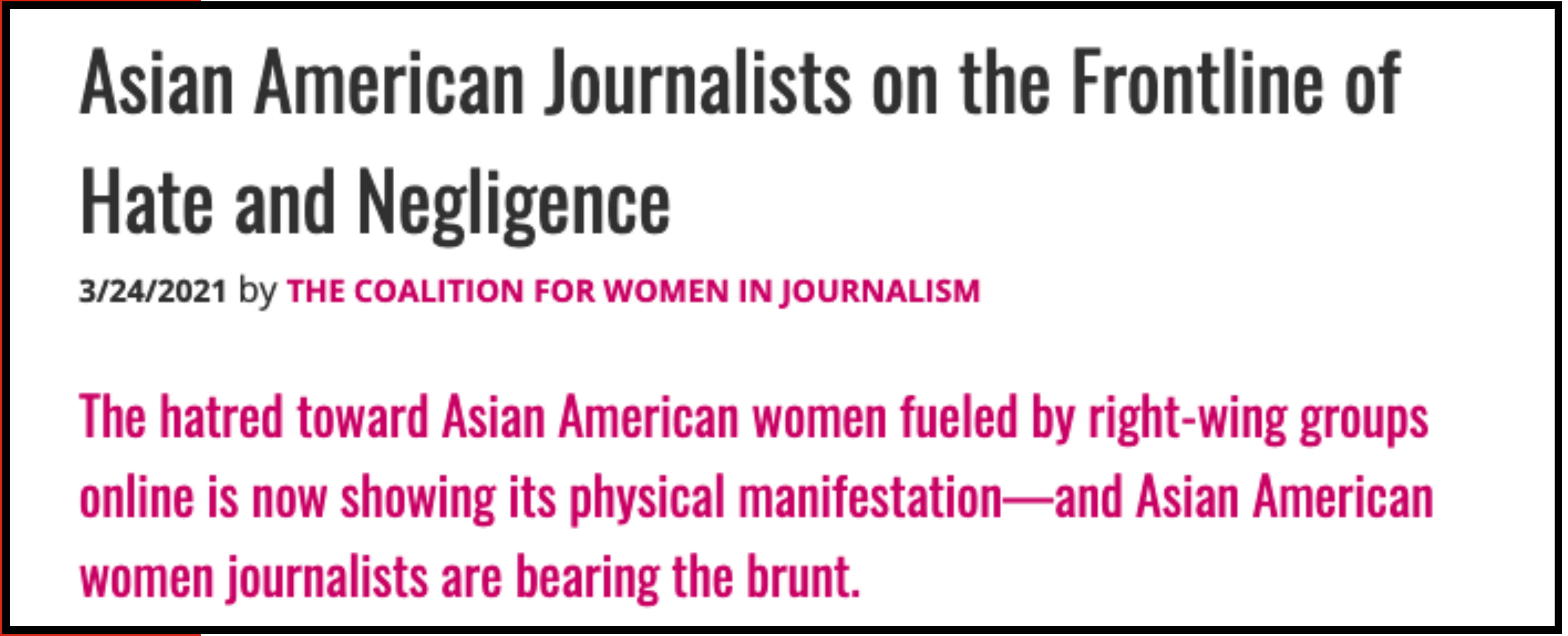


Online harassment of female journalists is real, and it's increasingly hard to endure

By Margaret Sullivan

Columnist

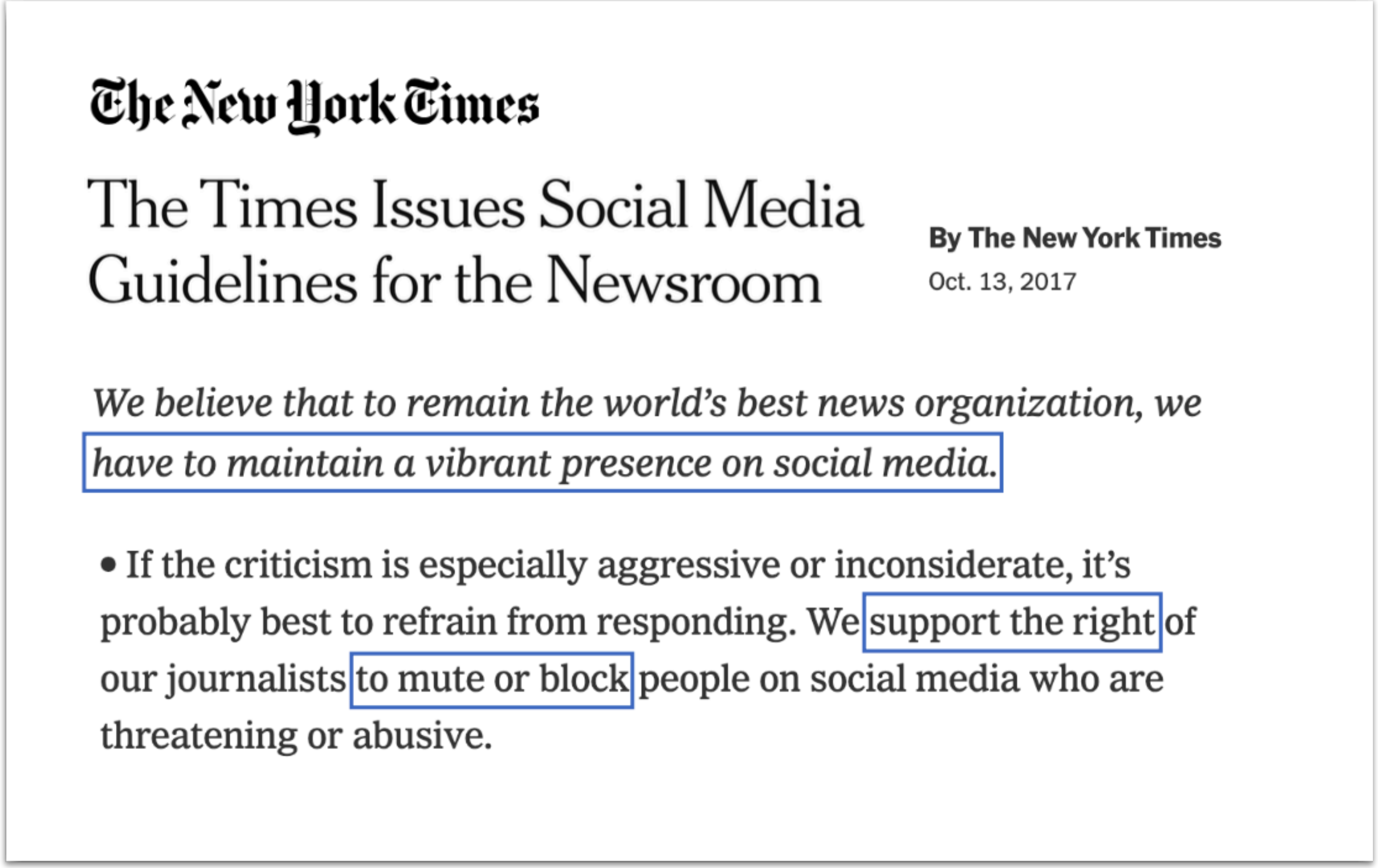
March 14, 2021



Asian American Journalists on the Frontline of Hate and Negligence

3/24/2021 by THE COALITION FOR WOMEN IN JOURNALISM

The hatred toward Asian American women fueled by right-wing groups online is now showing its physical manifestation—and Asian American women journalists are bearing the brunt.



The New York Times

The Times Issues Social Media Guidelines for the Newsroom

By The New York Times

Oct. 13, 2017

We believe that to remain the world's best news organization, we have to maintain a vibrant presence on social media.

- If the criticism is especially aggressive or inconsiderate, it's probably best to refrain from responding. We support the right of our journalists to mute or block people on social media who are threatening or abusive.

The things that made Twitter a powerful tool of social change were also the things that made it suck.

By: Sarah Jeong

Illustrations: Rui Pu & James Kerr

Dec 12, 2023, 06:00 AM PST

Twitter offers two tools to theoretically protect yourself [blocking and muting]… Since the platform indicates when you’ve been blocked by a user, the *Times* asked me not to do it to anyone… I wasn’t sure what was more unsettling: getting a death threat and seeing it, or not seeing it.

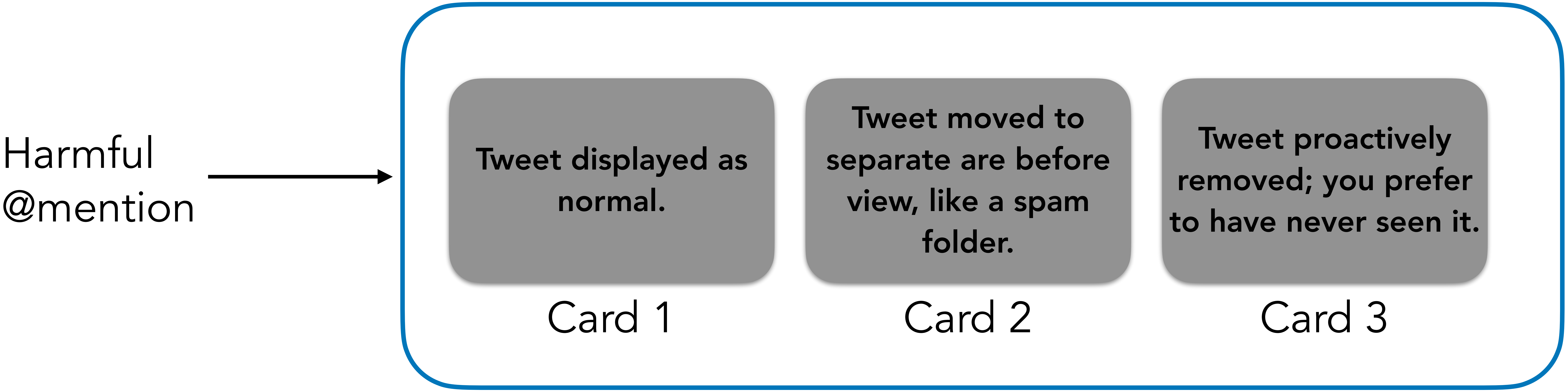
RQ1: How do journalists experience online harassment today and how do they protect themselves today?

RQ2: Would a defense informed by journalist norms be useful to journalists?

Need-finding Interviews

Study Design

- 8 participants (J1 – J8) from AAJA '22
  - Wide range of experience, newsroom size, and "beats"
  - Reply sorting: how would you process harmful replies?



Harmful @mention

Card 1: Tweet displayed as normal.

Card 2: Tweet moved to separate are before view, like a spam folder.

Card 3: Tweet proactively removed; you prefer to have never seen it.

Social media is a high value asset to journalists

Integral to daily life

- Every journalist we spoke to used social media, primarily Twitter/X, Instagram, Facebook
- Journalists use social media to source news stories, to engage with readership, to disseminate ideas, and participate in *online discourse*
  - Engage in *reciprocal journalism:* ***a mutually beneficial exchange between the journalist and their audiences***

Online engagement is necessary for the job

Integral to professionalism

Silencing themselves of dissenting readers viewed as **at odds** with their sense of duty:

> "If people want to criticize me, it still is in the public interest for people to be able to see my work. **People should be able to follow me and look at my tweets**, because I'm a journalist, and **I serve the public**." (J2)

Engagement is necessary *even* when harassed

Thresholds for engagement

3 participants stated they were more willing to interact if the exchange was related to their work – even if offensive or rude

6 participants opted to **defer** reading rather than remove harassing mentions to:

- Better understand their audience
- Monitor engagement for escalating abuse

Mitigation strategies

Limitations & Pitfalls

Blocking is **costly** and can **trigger further harassment.**

> “If feels like you are. giving something up when you choose to block somebody, and they can see that they got under your skin. I worry it would cause other people to [harass you] too.” (J8)

Summarizing journalists' needs

Utility and Reciprocation

- Use social media to contact sources, receive tips, gauge reader feedback
- Actively participate on social media to engage in *reciprocal journalism* and adhere to professional norms

Abuse-resistant Protections

- Mechanisms to stay apprised of harassment, especially during high-volume moments of harassment
- Use protections that don't validate attackers or trigger more harassment

Designing PressProtect

An anti-harassment system informed by journalists' needs

Designing PressProtect

An anti-harassment system informed by journalists' needs

Is this content harmful?

Is this content relevant?

Designing PressProtect

An anti-harassment system informed by journalists' needs

Is this content harmful?

JIGSAW

Leverage off-the-shelf harassment filters to identify harassment

Is this content relevant?

"All the News That's Fit to Print"

The New York Times

Late Edition

Today, cloudy, breezy, milder, rain tapering off, high 58. **Tonight,** clearing, low 37. **Tomorrow,** partly to mostly sunny, breezy, cooler, high 42. Weather map appears on Page B12.

VOL. CLXVII ... No. 57,851 © 2018 The New York Times Company NEW YORK, TUESDAY, JANUARY 23, 2018 $3.00

U.S. Watching While 2 Allies Clash in Syria

Turks' Attack on Kurds Upsets ISIS Fight

By MARK LANDLER and CARLOTTA GALL

WASHINGTON — When President Trump met with Turkey's president, Recep Tayyip Erdogan, at the United Nations last September, he embraced him as a friend and declared, "We're as close as we've ever been." Five months later, Turkey is waging an all-out assault against Syrian Kurds, America's closest allies in the war against the Islamic State.

The Turkish offensive, carried out over the protests of the United States but with the apparent assent of Russia, marks a perilous new phase in relations between two NATO allies — bringing their interests into direct conflict on the battlefield. It lays bare how much leverage the United States has lost in Syria, where its single-minded focus has been on vanquishing Islamist militants.

As Turkish troops advanced Monday on the Kurdish town of Afrin, in northwest Syria, the White House warned Turkey not to take its eye off the campaign against the Islamic State. But it stopped short of rebuking Turkey, and acknowledged its security concerns about the Kurds, whom Turkey considers terrorists and a threat to its territorial sovereignty.

The inherent conflict of the United States using the Kurds as its on-the-ground partner in fighting the Islamic State could be overlooked as long as that group remained a threat. But with the militants now in retreat, the White House is groping for a way to maintain relations with the Kurdish fighters without further alienating the Turks.

The Trump administration's response has been to help the Kurds build a border security force in northeast Syria, ostensibly to guard against the resurgence of the Islamic State. But that has only antagonized the Turks, who view it as a staging ground for a future insurgency against their homeland.

"The U.S. has tried to walk a very fine line in Syria," said Ali Soufan, a former F.B.I. counterterrorism agent who is now chairman of the Soufan Group. But

GABRIELLA DEMCZUK FOR THE NEW YORK TIMES

JOSHUA ROBERTS/REUTERS

WIN McNAMEE/GETTY IMAGES

House members leaving the Capitol on Monday, top. Senator Mitch McConnell, right, promised Senator Chuck Schumer, left, and fellow Democrats that he would allow an immigration vote.

Tax-Law Bonuses Veil a Bonanza for Companies

By JIM TANKERSLEY

WASHINGTON — There are good ways to start measuring how much the Trump tax cuts might be helping American workers. Tracking the bonus announcements flowing from corporations is not one of them.

Those announcements, which include $2,000 in stock grants for

$1,000 bonuses for Bank of America employees, are both real money and smart marketing. President Trump and top Republicans lawmakers have praised many of the companies that are disclosing tax-cut-fueled bonuses and wage hikes.

For the most part, though, they are not indicative of the windfalls that

years to come.

Companies are acknowledging this in their fourth-quarter earnings reports and other financial disclosures, which earmark just a sliver of their future tax savings for direct and indirect investments in workers.

Bank of America's bonuses will cost the bank $145 million in 2018, or about 5 percent of the nearly

SHUTDOWN ENDS, SETTING UP CLASH OVER 'DREAMERS'

NEWS ANALYSIS

Hope for Bargain in a Swift Surrender

By CARL HULSE

WASHINGTON — Senate Democrats blinked. But the saving grace for them may be that they did it quickly.

"Should we have ever shut down the government?" asked Senator Joe Manchin III of West Virginia, who played a leading role in finding a way out for Democrats. "Absolutely not."

Over the weekend it became clear that using the shutdown to insist on protections for hundreds of thousands of undocumented immigrants was a serious miscalculation. By abandoning the strategy on its third day, Democrats believe they limited whatever damage there may be and gave the public time to forget about the disruption before the crucial November election.

They also won a commitment from Senator Mitch McConnell, the Kentucky Republican and majority leader, to allow debate on a wide-ranging, bipartisan measure that could achieve their goal of protecting hundreds of thousands of undocumented immigrants.

But by supporting the government's reopening, they provoked a surge of outrage from the party's left as progressive activists and lawmakers assailed the deal as a capitulation based on a mere promise by Mr. McConnell, a longtime foe known for his obstruction of the Democratic agenda.

Immigration activists were furious that Senator Chuck Schumer of New York, the Democratic leader, and his fellow Democrats had given away the best leverage they had to ensure that the young immigrants were protected.

"It's morally reprehensible and it's political malpractice," said Ezra Levin, the co-executive director of Indivisible, a liberal group.

"Leader McConnell's track record of keeping his

Immigration Issues Unresolved Under Senate Accord

By SHERYL GAY STOLBERG and THOMAS KAPLAN

WASHINGTON — Congress brought an end to a three-day government shutdown on Monday as Senate Democrats buckled under pressure to adopt a short-term spending bill to fund government operations without first addressing the fate of young undocumented immigrants.

The House quickly approved the measure — which will fund the government through Feb. 8 and extend funding for the popular Children's Health Insurance Program for six years — and President Trump signed it on Monday night.

The agreement also revealed fissures among Democrats, with about one-third of the party's members in the Senate and a majority in the House voting against it.

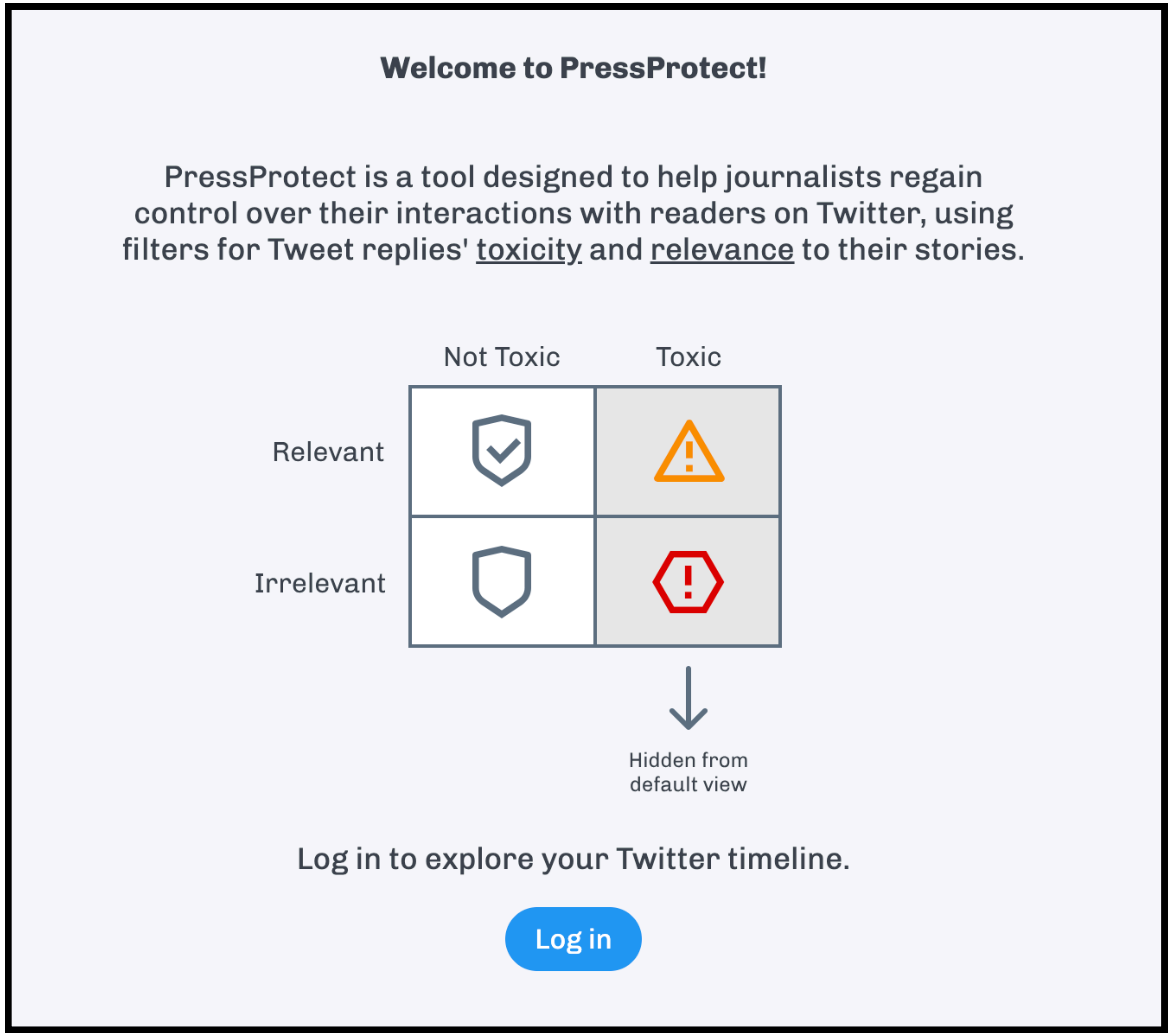
The passage of the measure ended an ugly, if short-lived, impasse that threatened to give a black eye to both major political parties. The deal, reached after a bipartisan group of senators pushed their leaders to come to terms, enables hundreds of thousands of federal employees who had been facing furloughs to go back to work.

But a key part of the deal, a pledge by Senator Mitch McConnell, Republican of Kentucky and the majority leader, to allow an immigration vote in the coming weeks, sets the stage for a battle over the so-called Dreamers, young undocumented immigrants who were brought to the United States illegally as children.

Lawmakers in the House and Senate are offering drastically different visions of how to resolve their fate. But those on both sides of the debate, as well as advocates for immigrants' rights, said that ultimately Mr. Trump would need to get involved for the immigration dispute to be settled.

Relevance algorithms to identify if incoming messages are related to a given article

PressProtect adds **client-side UI controls** for viewing *harmful* content.



Welcome to PressProtect!

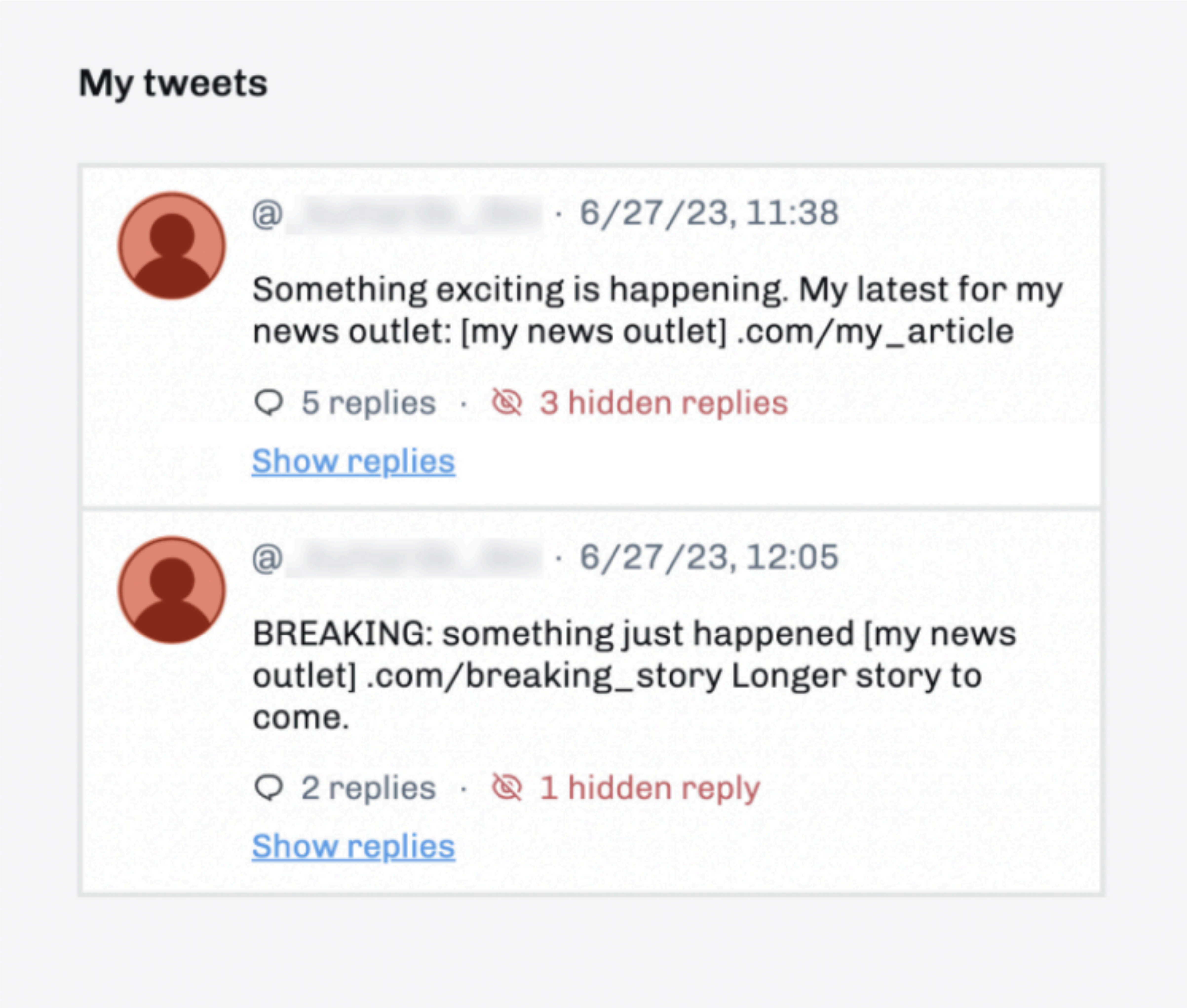
PressProtect is a tool designed to help journalists regain control over their interactions with readers on Twitter, using filters for Tweet replies' toxicity and relevance to their stories.

	Not Toxic	Toxic
Relevant		
Irrelevant		

Hidden from default view

Log in to explore your Twitter timeline.

UI: Home Page



My tweets

@[redacted] · 6/27/23, 11:38

Something exciting is happening. My latest for my news outlet: [my news outlet] .com/my_article

5 replies · 3 hidden replies

Show replies

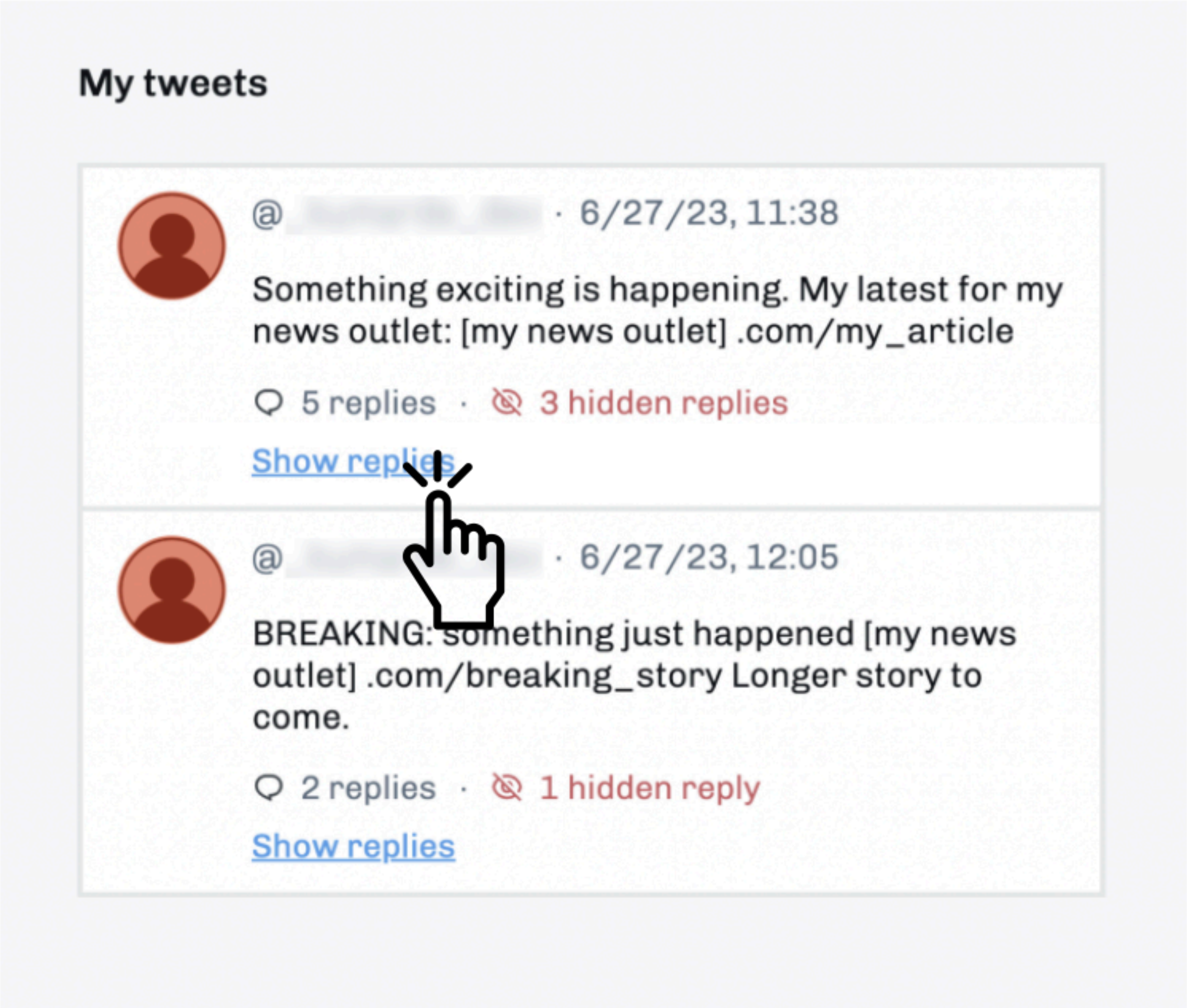
@[redacted] · 6/27/23, 12:05

BREAKING: something just happened [my news outlet] .com/breaking_story Longer story to come.

2 replies · 1 hidden reply

Show replies

UI: Home Page



My tweets

@[redacted] · 6/27/23, 11:38

Something exciting is happening. My latest for my news outlet: [my news outlet] .com/my_article

5 replies · 3 hidden replies

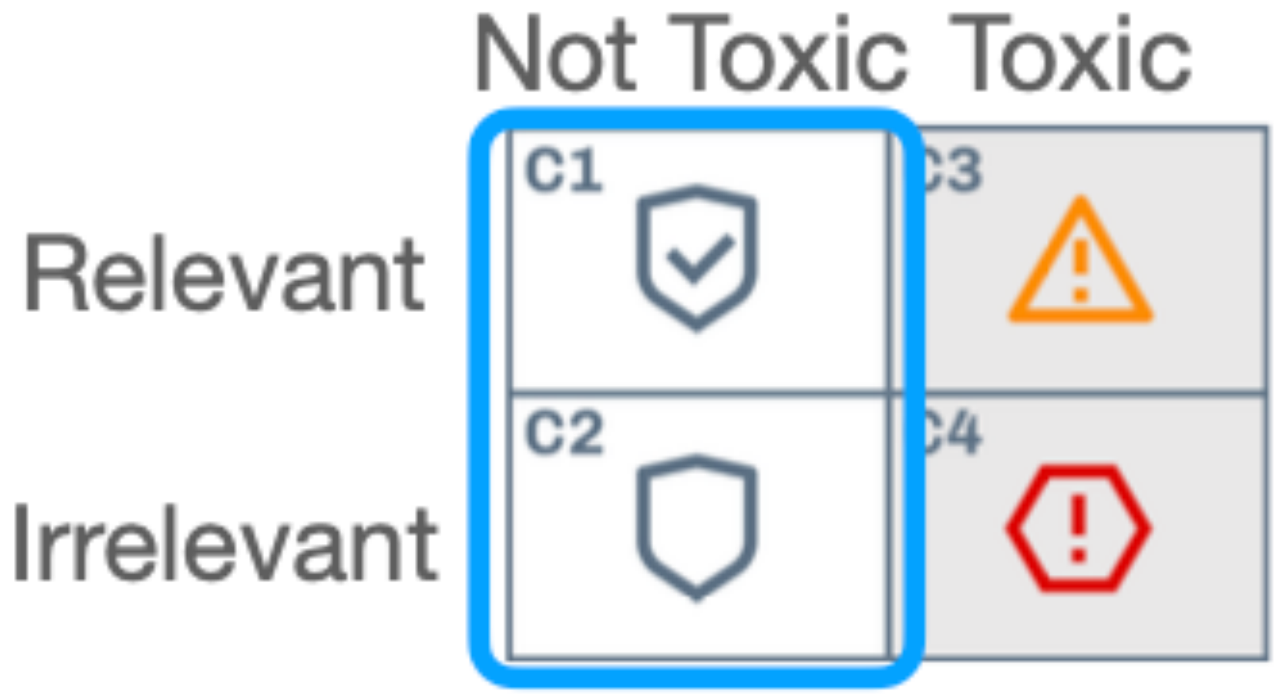
Show replies

@[redacted] · 6/27/23, 12:05

BREAKING: something just happened [my news outlet] .com/breaking_story Longer story to come.

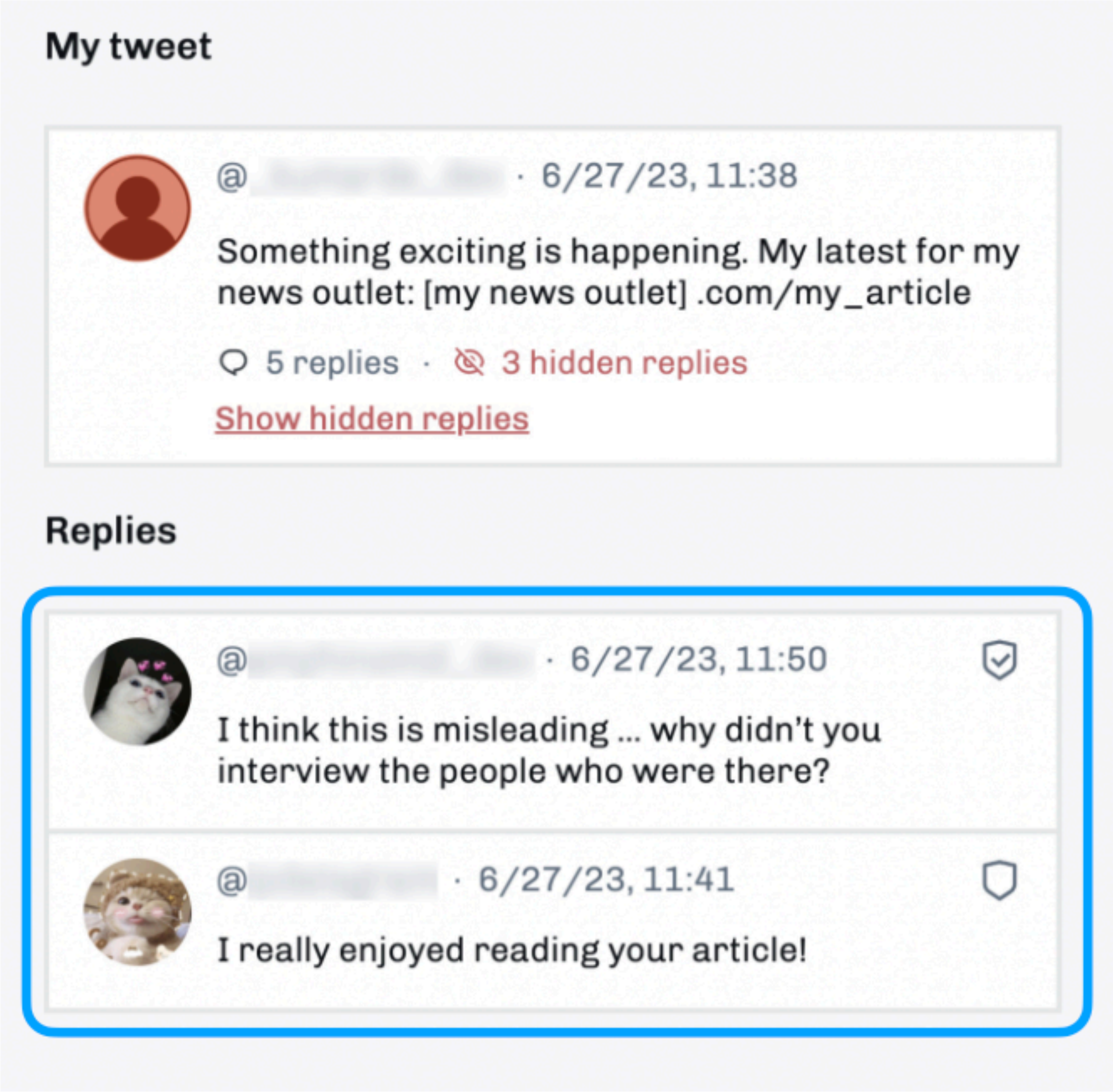
2 replies · 1 hidden reply

Show replies



	Not Toxic	Toxic
Relevant	C1	C3
Irrelevant	C2	C4

UI: Tweet replies (default)



My tweet

@[redacted] · 6/27/23, 11:38

Something exciting is happening. My latest for my news outlet: [my news outlet] .com/my_article

5 replies · 3 hidden replies

Show hidden replies

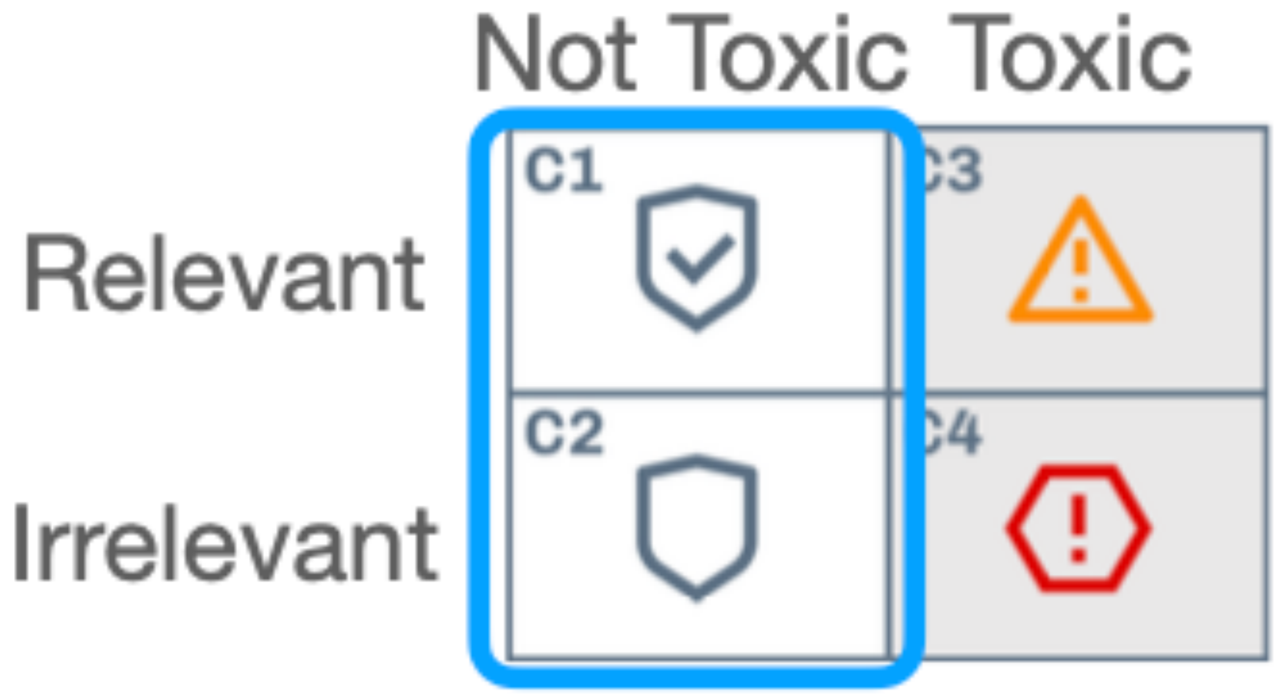
Replies

@[redacted] · 6/27/23, 11:50

I think this is misleading ... why didn't you interview the people who were there?

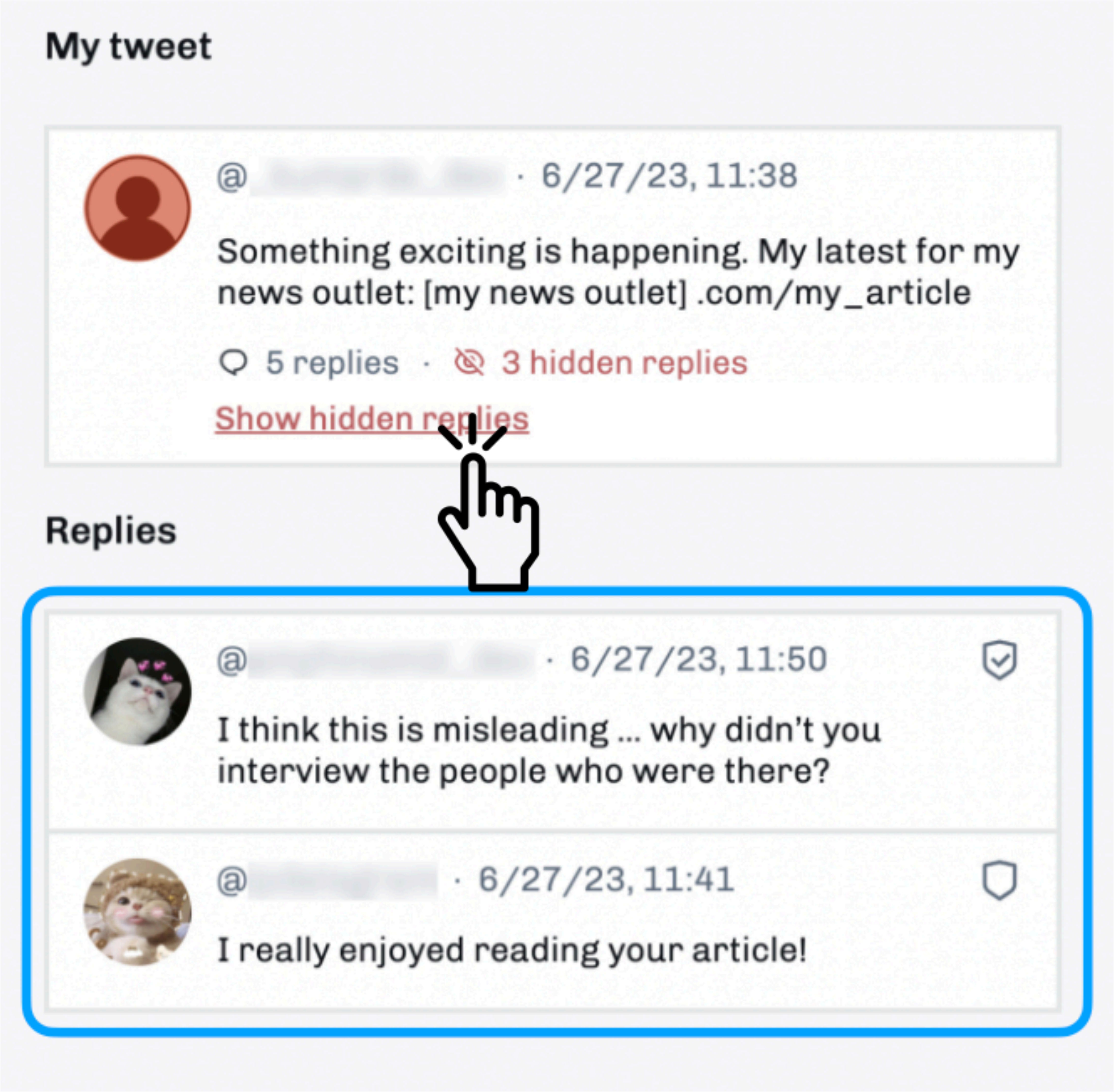
@[redacted] · 6/27/23, 11:41

I really enjoyed reading your article!



	Not Toxic	Toxic
Relevant	C1	C3
Irrelevant	C2	C4

UI: Tweet replies (default)



My tweet

@[redacted] · 6/27/23, 11:38

Something exciting is happening. My latest for my news outlet: [my news outlet] .com/my_article

5 replies · 3 hidden replies

Show hidden replies

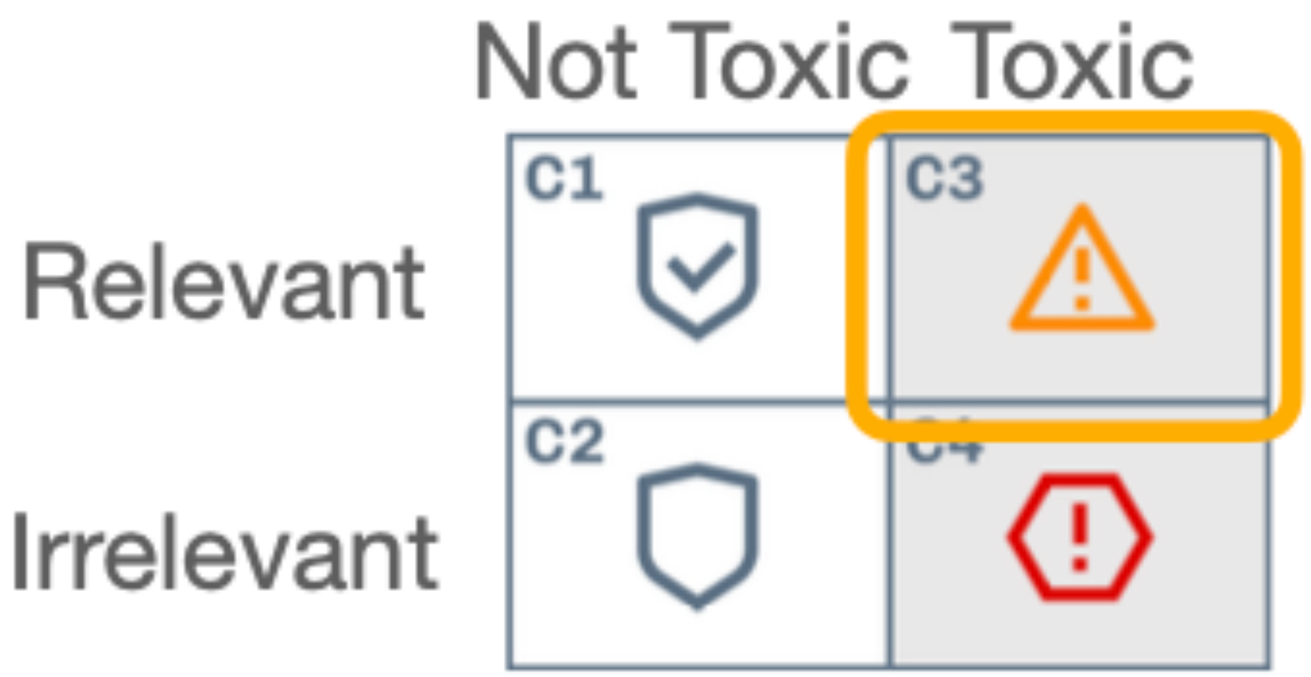
Replies

@[redacted] · 6/27/23, 11:50

I think this is misleading ... why didn't you interview the people who were there?

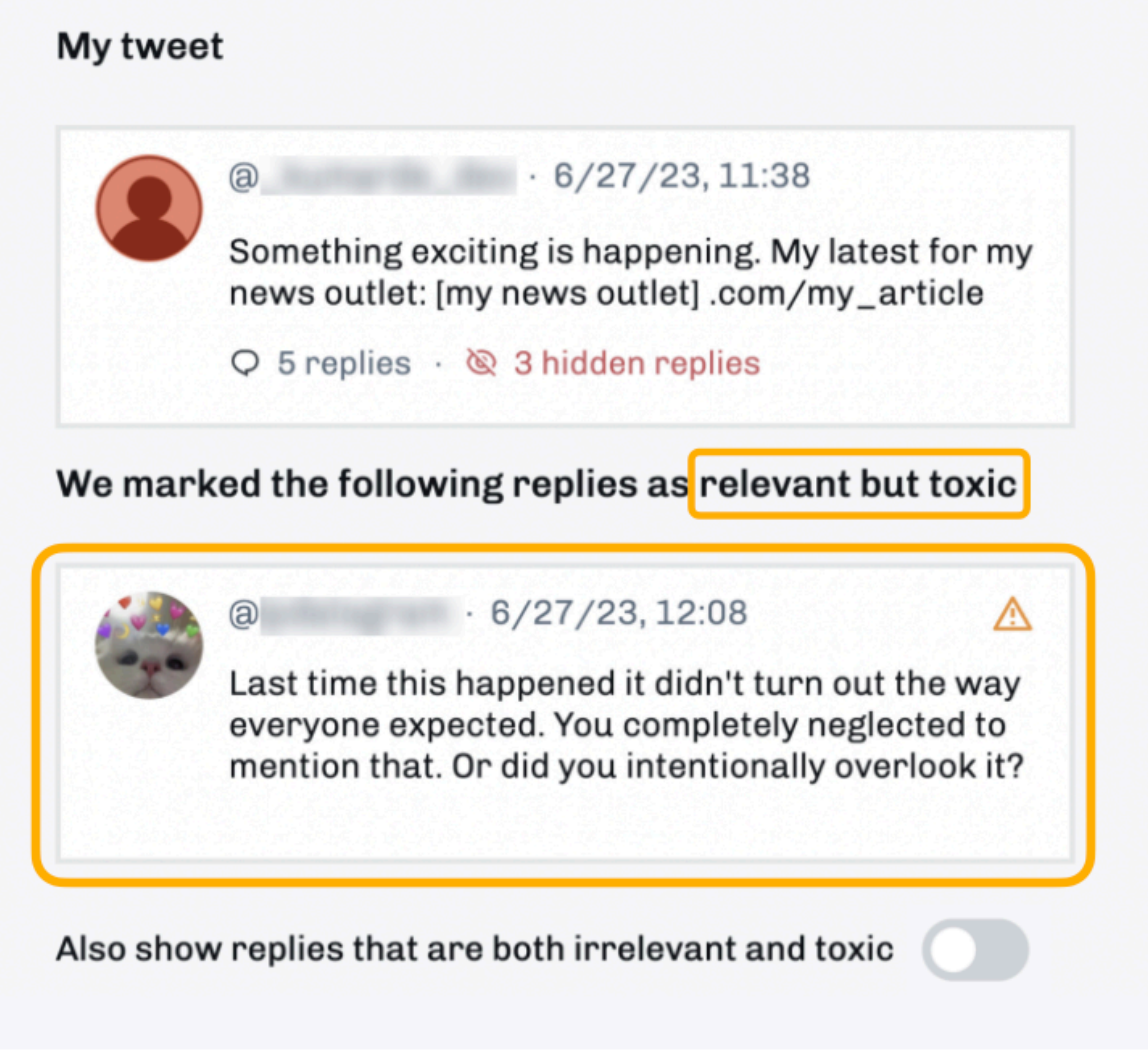
@[redacted] · 6/27/23, 11:41

I really enjoyed reading your article!



	Not Toxic	Toxic
Relevant	C1	C3
Irrelevant	C2	C4

UI: Tweet replies (harmful)



My tweet

@[redacted] · 6/27/23, 11:38

Something exciting is happening. My latest for my news outlet: [my news outlet] .com/my_article

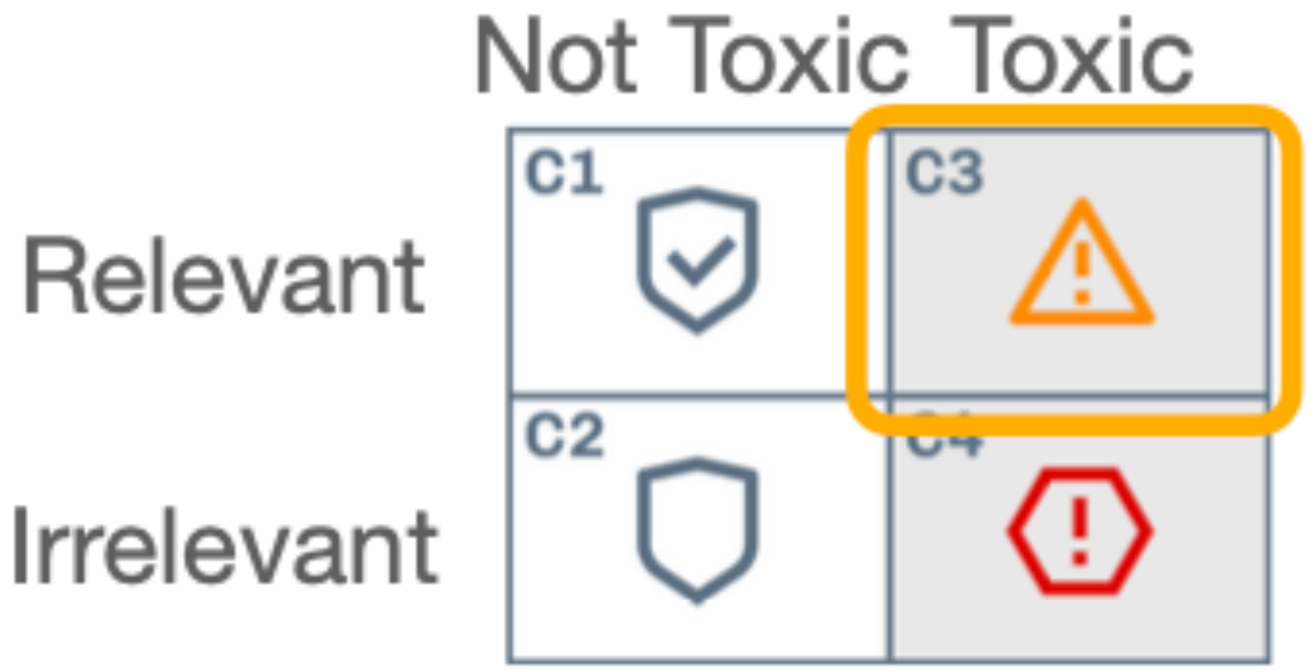
5 replies · 3 hidden replies

We marked the following replies as relevant but toxic

@[redacted] · 6/27/23, 12:08 ⚠

Last time this happened it didn't turn out the way everyone expected. You completely neglected to mention that. Or did you intentionally overlook it?

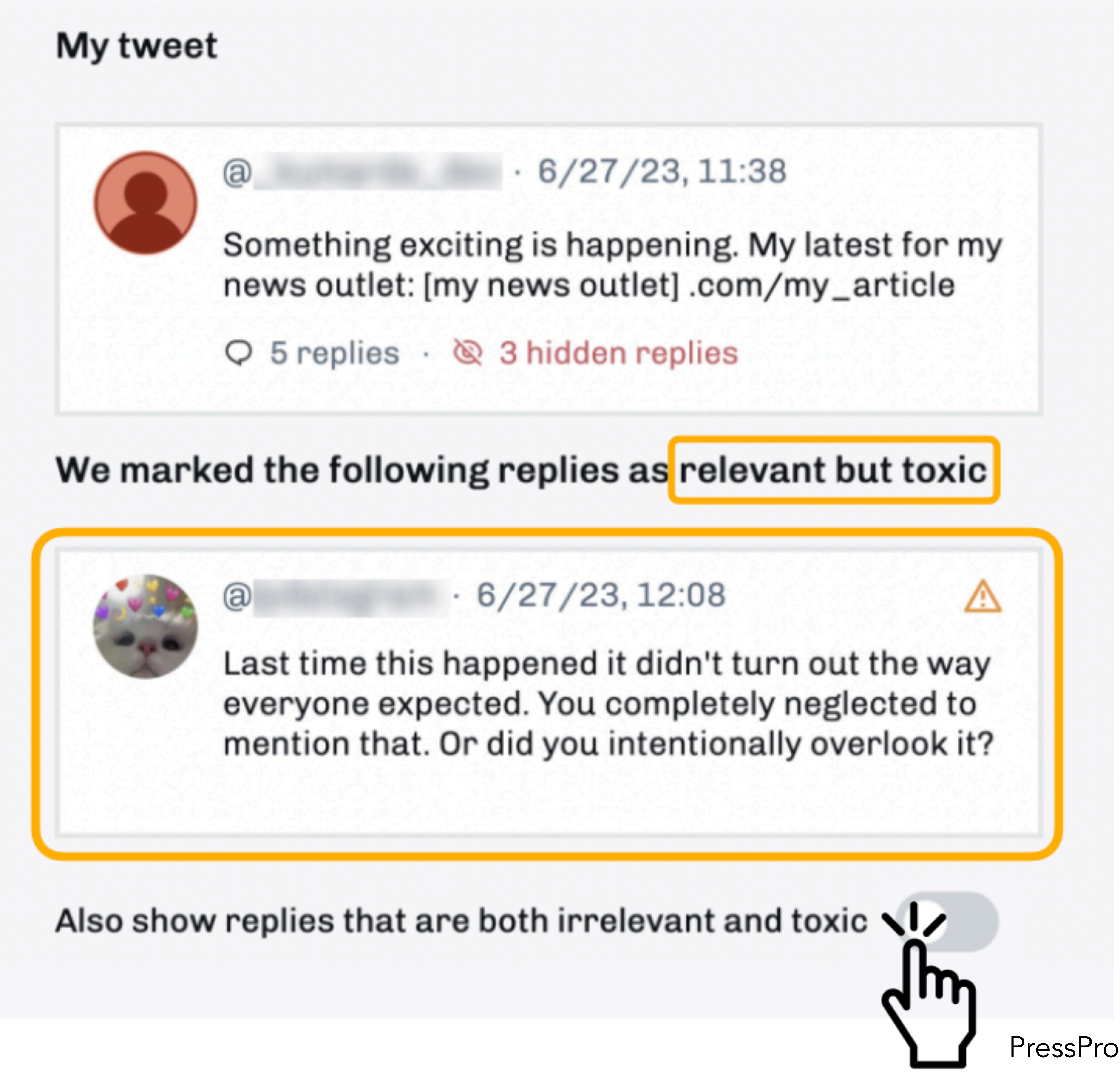
Also show replies that are both irrelevant and toxic ○



	Not Toxic	Toxic
Relevant	C1 ✓	C3 ⚠
Irrelevant	C2	C4 !

PressProtect: Helping Journalists Navigate Social Media in the Face of Online Harassment

UI: Tweet replies (harmful)



My tweet

@[redacted] · 6/27/23, 11:38

Something exciting is happening. My latest for my news outlet: [my news outlet] .com/my_article

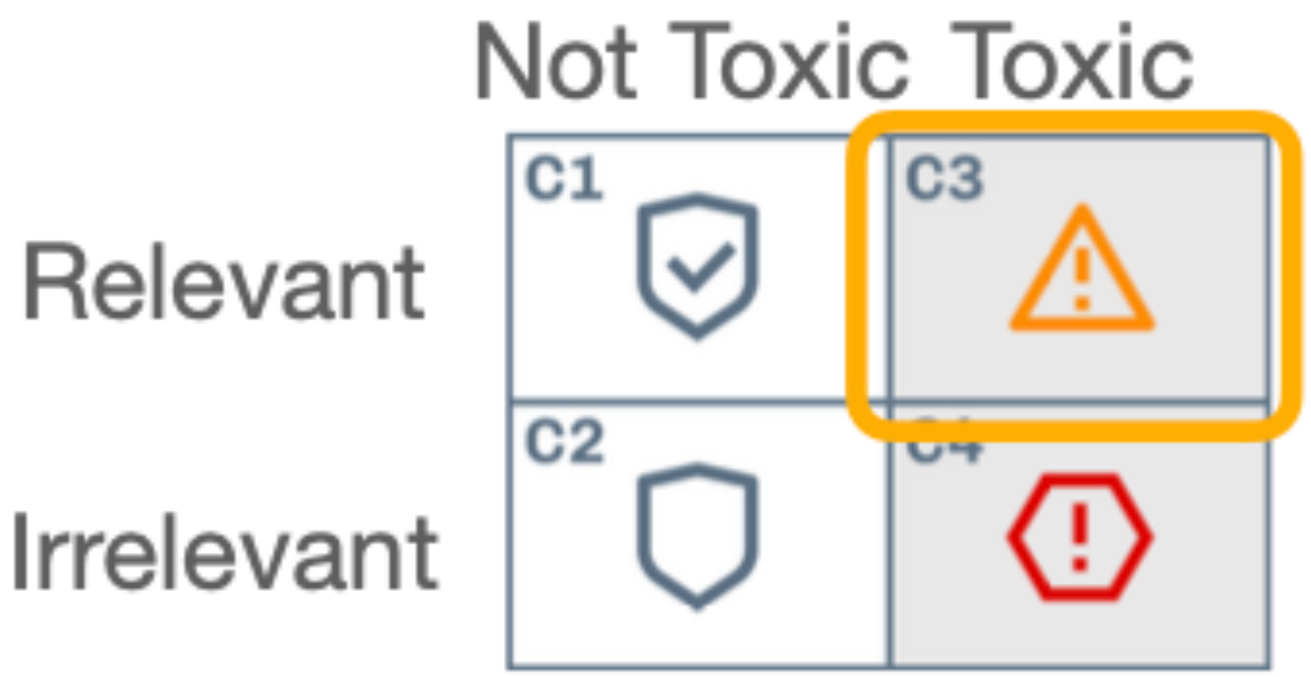
5 replies · 3 hidden replies

We marked the following replies as relevant but toxic

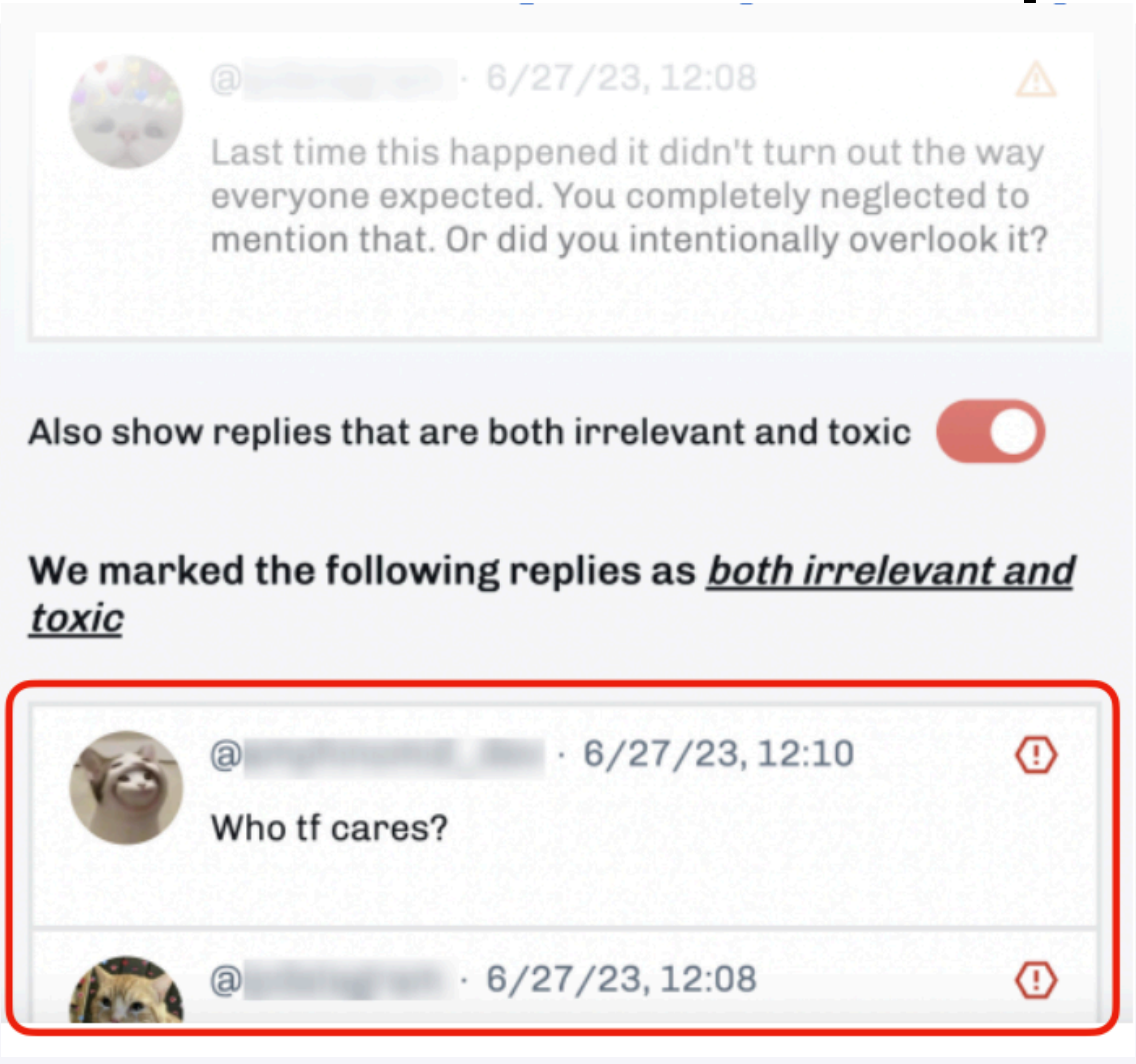
@[redacted] · 6/27/23, 12:08 ⚠

Last time this happened it didn't turn out the way everyone expected. You completely neglected to mention that. Or did you intentionally overlook it?

Also show replies that are both irrelevant and toxic



UI: Tweet replies (harmful)



@[redacted] · 6/27/23, 12:08

Last time this happened it didn't turn out the way everyone expected. You completely neglected to mention that. Or did you intentionally overlook it?

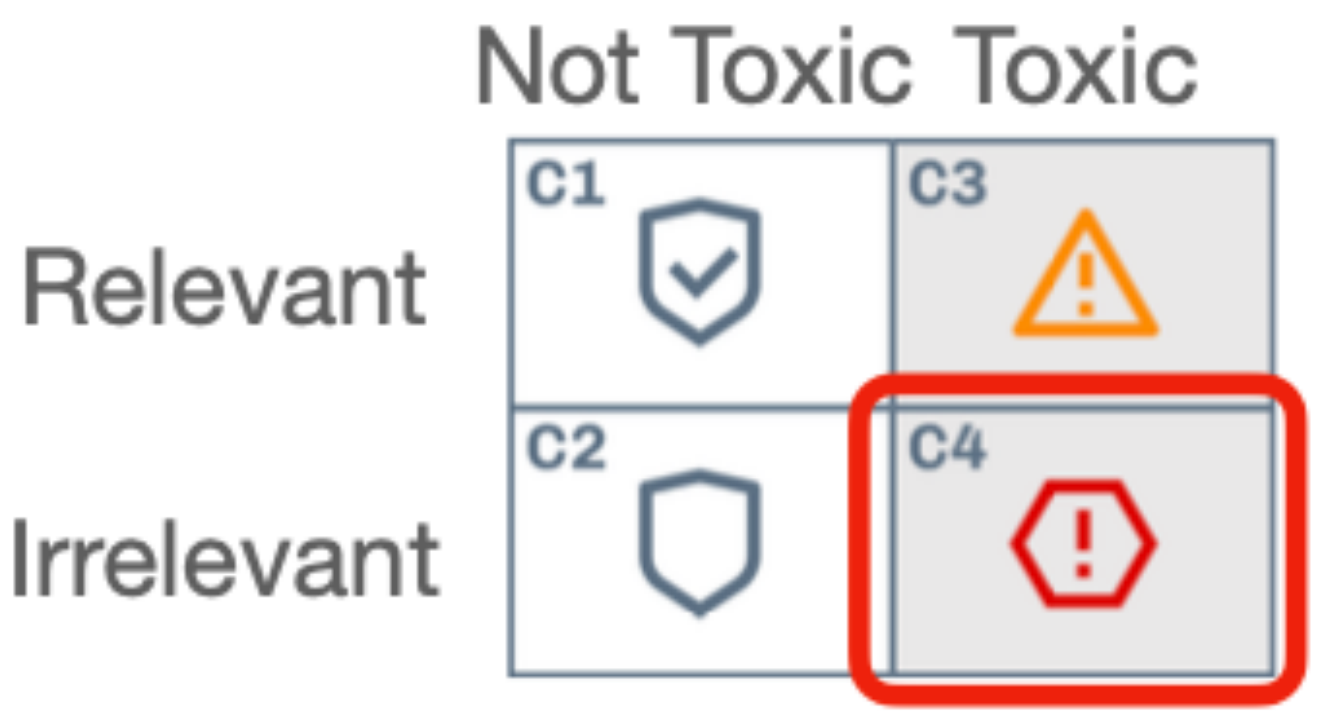
Also show replies that are both irrelevant and toxic

We marked the following replies as ***both irrelevant and toxic***

@[redacted] · 6/27/23, 12:10

Who tf cares?

@[redacted] · 6/27/23, 12:08



	Not Toxic	Toxic
Relevant	C1	C3
Irrelevant	C2	C4

PressProtect: Helping Journalists Navigate Social Media in the Face of Online Harassment

Evaluating PressProtect

User testing for PressProtect

- 8 journalists (P1 – P8) who were active on Twitter and authored tweets promoting their stories that received replies
  - 2 participants from initial need-finding interviews
  - 6 others from cold-emailing and snowball sampling

What did journalists think about the system?

Exit-interviews with journalists that used the tool on their own profiles

- *Participants felt PressProtect protected them against harassment that could generalize to serve other visibly online users*
- *Participants value the ability to customize automated tools for a wide range of personal preferences*
- *Participants did not feel PressProtect would hinder valuable interactions with readers – satisfied with the tradeoff for protection*

Participants wanted a distinction between imminent physical threats and other harassment.

- 5 participants wanted such threats to be flagged
- 1 participant was *critically* concerned with PressProtect **obscuring threats.**
- Monitoring preferences are shaped by risk perception:

> "I would just stop using [PressProtect] because of **my bias towards wanting to see every thing…** I just want to, as some point, have seen them all… You still really need to **monitor** what people are saying to make sure that it doesn't translate into **physical danger**" (P7)

Limitations of PressProtect

- Real-time was impossible to do with API changes @ Twitter back in 2023 / 2024
    - Became prohibitively expensive to do this, but opportunities exist in Fediverse / decentralized style platforms
- Relies on a lot of external tooling
    - To deploy this in practice, we'd need to minify / run models locally
- Privacy is a **big** concern
    - Users shouldn't have to have all their interactions screened by an external entity to gain protections

What did we learn about community defenses?

- Anti-harassment needs are **multi-dimensional** and shaped by the community: simply filtering content with machine learning does **not** work for everyone
- Abstraction was effective and users appreciated control —> but there is an emergent need to identify “just online” from “online and maybe real-world” threats
- Participants felt tool could be generalizable —> that’s next!

Ongoing projects…

- Building systems to identify AI slop citations in federal legal proceedings
- Exploring content moderation practices on Bluesky and the third party moderation ecosystem
- Breaking content authenticity protocols for AI generated content
- Studying how image resharing and image manipulation shapes news media
- Studying how AI generated content on YouTube is recommended to people
- Studying how ad targeting works on mobile apps (TikTok) and how income plays a role in targeted advertising
- Exploring how to discover new misinformation domains before they have the ability to spread quickly and launch narratives…

For more research, see https://kumarde.com/publications

The end!

- Thanks so much for being in the class. You’ve been awesome and I hope it’s been fun!
- If you’re interested in more security classes…
    - I’m teaching CSE 227 in the Spring (graduate computer security, but it’s a lot more research-y)
    - I’m teaching a CSE190 in the Fall about web tracking and web privacy, if you’re still on campus!
- If you’re interested in seeing a play “about” cybersecurity…
    - https://www.scr.org/plays/productions/25-26-season/advanced-persistent-teenagers/