Discussion of Privacy & Security

Gabrielle Lake, Ann-Cathrin Lena Kloeckner, Weijun Li

Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks

Hao-Ping (Hank) Lee, Yu-Ju Yang, Thomas Serban Von Davier, Jodi Forlizzi, and Sauvik Das

CHI 2024; Best Paper

LLM
LARGE LANGUAGE MODEL

•••• ••••

https://www.menti.com/al1wd8ciqk4a

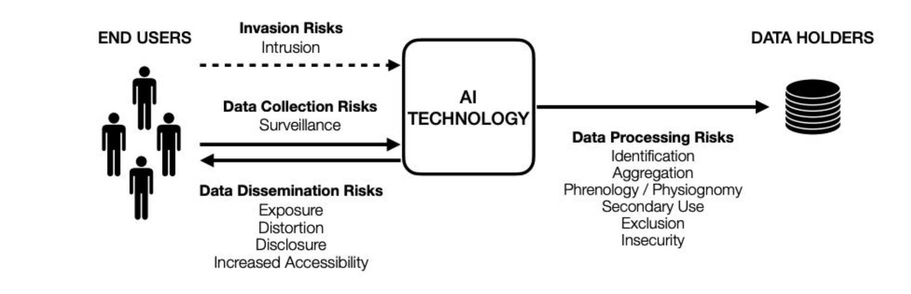
General Overview

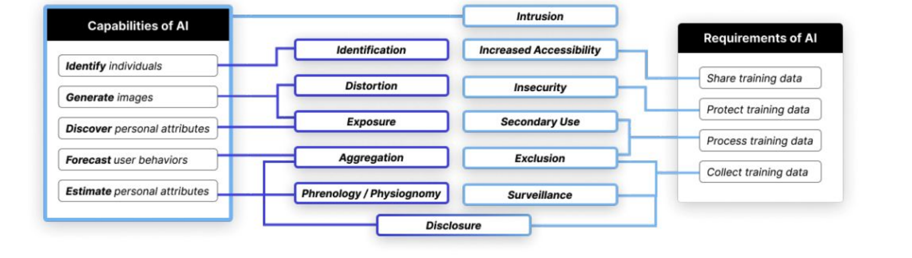
1️⃣How does AI change privacy risks

2️⃣Conducts an empirical analysis of 321 AI privacy incidents to develop a taxonomy of AI privacy risks(AIAAIC).

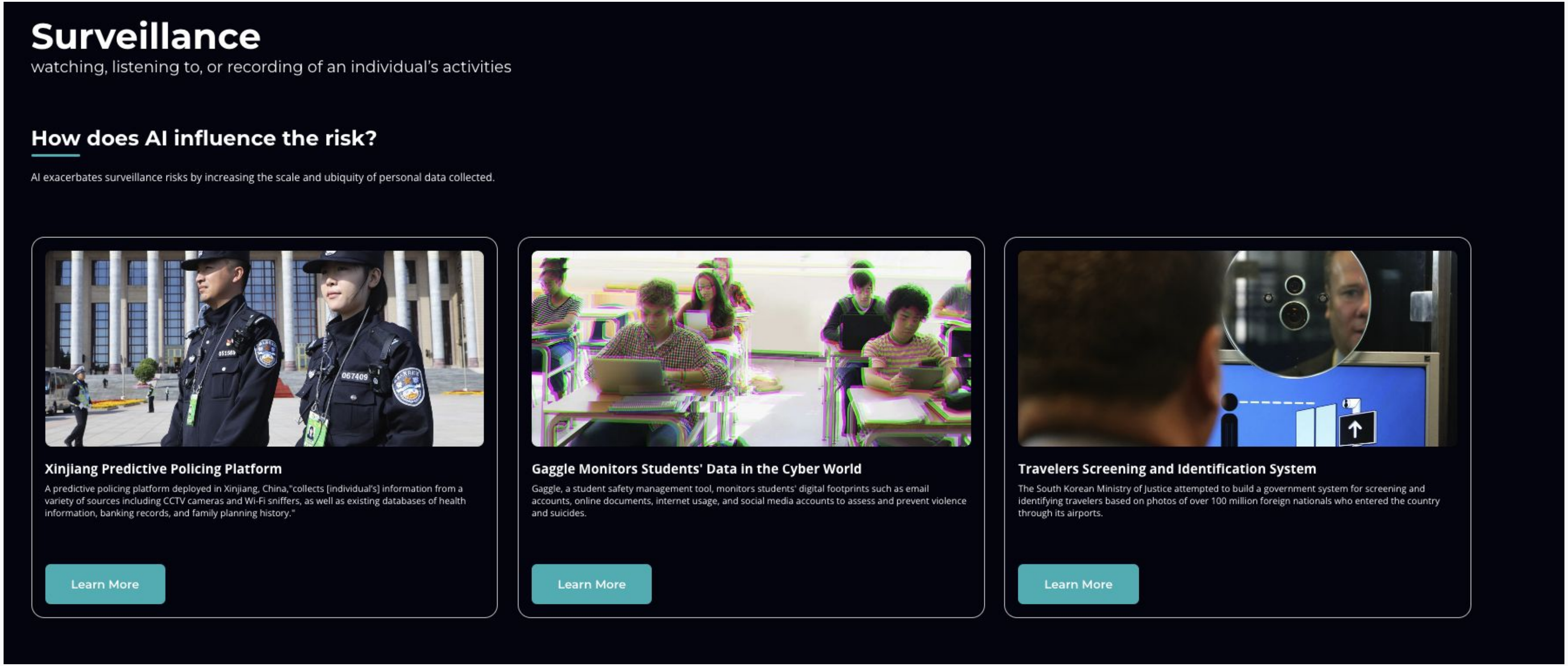
3️⃣Finds 12 privacy risks, categorized into four main areas: data collection, data processing, data dissemination, and invasion risks, and contextualizes them within existing privacy frameworks.

4️⃣Contributes recommendations for HCI researchers, policymakers, and AI developers to mitigate privacy risks and develop AI systems that prioritize user privacy and security.





General Overview - Data Collection



Surveillance

watching, listening to, or recording of an individual's activities

How does AI influence the risk?

AI exacerbates surveillance risks by increasing the scale and ubiquity of personal data collected.

Xinjiang Predictive Policing Platform

A predictive policing platform deployed in Xinjiang, China,"collects [individual's] information from a variety of sources including CCTV cameras and Wi-Fi sniffers, as well as existing databases of health information, banking records, and family planning history."

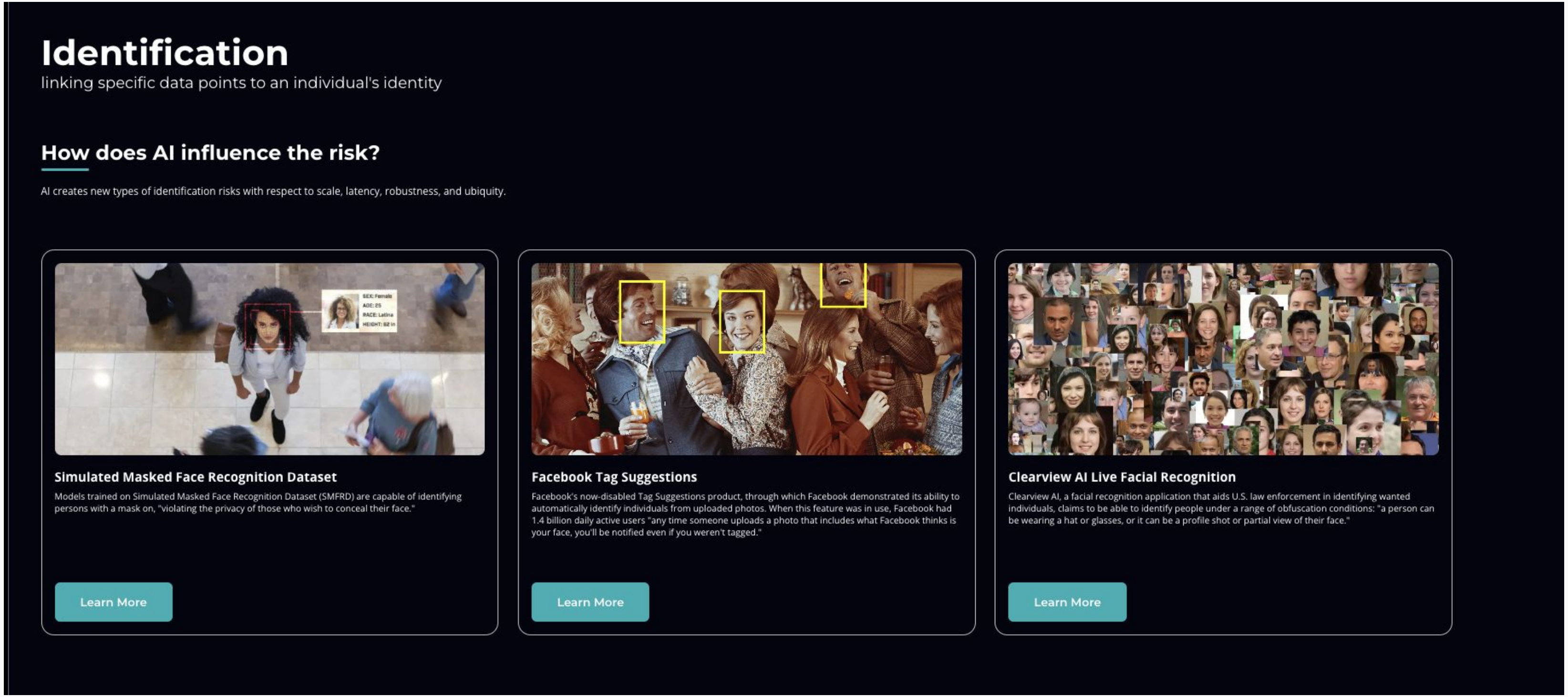
Gaggle Monitors Students' Data in the Cyber World

Gaggle, a student safety management tool, monitors students' digital footprints such as email accounts, online documents, internet usage, and social media accounts to assess and prevent violence and suicides.

Travelers Screening and Identification System

The South Korean Ministry of Justice attempted to build a government system for screening and identifying travelers based on photos of over 100 million foreign nationals who entered the country through its airports.

General Overview - Data Processing



Identification

linking specific data points to an individual's identity

How does AI influence the risk?

AI creates new types of identification risks with respect to scale, latency, robustness, and ubiquity.

Simulated Masked Face Recognition Dataset

Models trained on Simulated Masked Face Recognition Dataset (SMFRD) are capable of identifying persons with a mask on, "violating the privacy of those who wish to conceal their face."

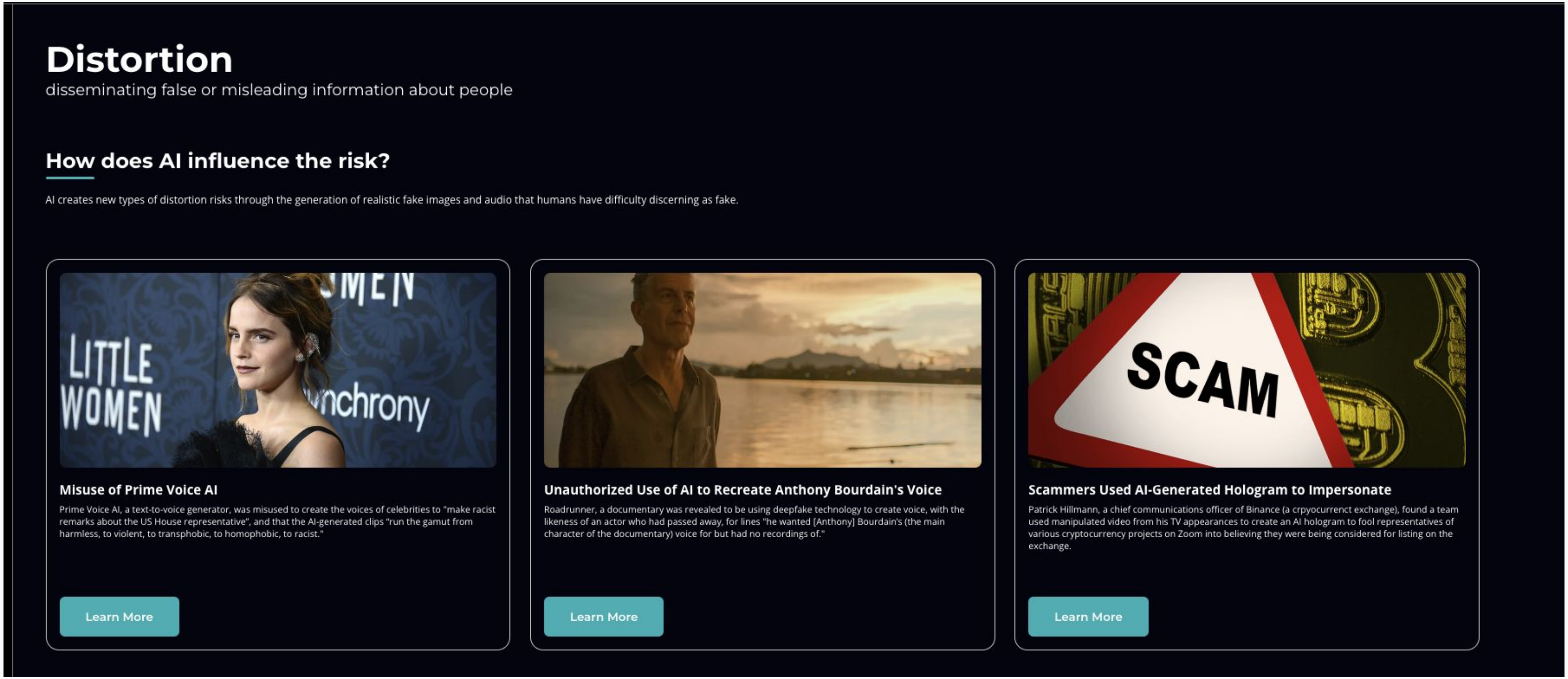
Facebook Tag Suggestions

Facebook's now-disabled Tag Suggestions product, through which Facebook demonstrated its ability to automatically identify individuals from uploaded photos. When this feature was in use, Facebook had 1.4 billion daily active users "any time someone uploads a photo that includes what Facebook thinks is your face, you'll be notified even if you weren't tagged."

Clearview AI Live Facial Recognition

Clearview AI, a facial recognition application that aids U.S. law enforcement in identifying wanted individuals, claims to be able to identify people under a range of obfuscation conditions: "a person can be wearing a hat or glasses, or it can be a profile shot or partial view of their face."

General Overview - Data Dissemination



Distortion

disseminating false or misleading information about people

How does AI influence the risk?

AI creates new types of distortion risks through the generation of realistic fake images and audio that humans have difficulty discerning as fake.

Misuse of Prime Voice AI

Prime Voice AI, a text-to-voice generator, was misused to create the voices of celebrities to "make racist remarks about the US House representative", and that the AI-generated clips "run the gamut from harmless, to violent, to transphobic, to homophobic, to racist."

Unauthorized Use of AI to Recreate Anthony Bourdain's Voice

Roadrunner, a documentary was revealed to be using deepfake technology to create voice, with the likeness of an actor who had passed away, for lines "he wanted [Anthony] Bourdain's (the main character of the documentary) voice for but had no recordings of."

Scammers Used AI-Generated Hologram to Impersonate

Patrick Hillmann, a chief communications officer of Binance (a crpyocurrenct exchange), found a team used manipulated video from his TV appearances to create an AI hologram to fool representatives of various cryptocurrency projects on Zoom into believing they were being considered for listing on the exchange.

Worthy of Study?

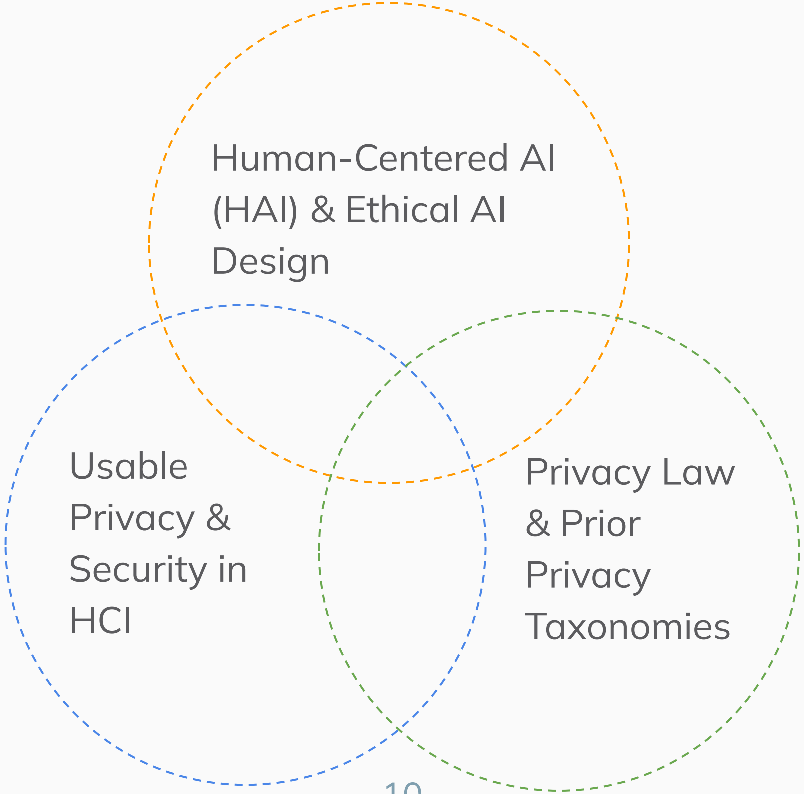
1️⃣ AI introduces new privacy risks

2️⃣Existing privacy frameworks do not fully capture AI-specific threats.

3️⃣ AI practitioners lack awareness and incentives to address these issues.

4️⃣Public concerns over AI privacy are growing, yet solutions remain limited

Areas of Inquiry



Human-Centered AI (HAI) & Ethical AI Design

Usable Privacy & Security in HCI

Privacy Law & Prior Privacy Taxonomies

HCI Research Contributions

1️⃣Identifying and Categorizing AI-Specific Privacy Risks

2️⃣Informing Human-Centered AI Design

3️⃣Bridging HCI, Privacy Law, and AI Ethics

💡 Prior HCI privacy research has focused on user perceptions of privacy and security, but this paper expands that scope by offering a structured way to evaluate AI-driven privacy risks. It provides both theoretical insights and practical tools for AI and HCI practitioners,

Main Takeaways

HCI Community

1️⃣Existing HCI frameworks are insufficient to address these challenges, and the AI Privacy Risk Taxonomy introduced in this paper provides a structured tool for analyzing and designing privacy-preserving AI systems.

2️⃣Empirical, **incident-Based** analysis is critical for AI privacy research

3️⃣HCI needs to rethink privacy design in AI-powered products

4️⃣Calls for new transparency and risk-assessment tools in HCI to help users and developers better understand AI's privacy implications.

For AI Research

The AI Privacy Risk Taxonomy provides a structured framework for incorporating privacy considerations into AI development.

For Privacy Research

Expand their focus beyond protecting stored data to prevent AI from generating or inferring private information that users never directly provided.

For Law & Policy Research

Explore new privacy protection strategies that address AI-specific concerns

Discussion questions

Which of the new privacy risks identified in the taxonomy do you find most concerning, and why? How might these risks evolve as AI technologies continue to advance?

How can AI practitioners effectively use this taxonomy to identify and mitigate potential privacy risks in their products or services? What challenges might they face in doing so?

Analyzing Security and Privacy Advice During the 2022 Russian Invasion of Ukraine on Twitter

Juliane Schmuser, Harshini Sri Ramulu, Noah Wohler, Christian Stransky, Felix Bensmann, Dimitar Dimitrov, Sebastian Schellhammer, Dominik Wermke, Stefan Dietze, Yasemin Acar, and Sascha Fahl

CHI 2024; Honorable Mention

Discussion

https://www.menti.com/alsd6jekd9hr

General Overview

Introduces the 2022 Russian invasion of the Ukraine as an initiator for an evolving cyber threat landscape.

Explains that social media became a key platform for disseminating security and privacy advice.

Analyzes advice from tweets and linked documents.

Focuses on effectiveness of advice, misinformation, and emerging security trends.

Aims to identify patterns, effectiveness, and gaps in security recommendations.

Security Threats in Cyberspace During Crisis

Misinformation: Distinguishing between genuine advice and misinformation and genuine threats and wrong reports is important to ensure there can be an effective crisis communication.

Disinformation / exploitation: Spreading unverified or malicious privacy tools as “secure solutions”. This puts users at risk.

Erosion of trust and confusion: State-sponsored misinformation campaigns, eroding trust in governments.

Rapid development: Constantly changing environments - Imminent response especially important in conflict environments.

Why Is This Relevant?

Comparison to previous work on security advice

- Similar focus on strong passwords, multi-factor authentication, and software updates
- Social media communication and information sharing

Innovation: Crisis-specific recommendations (e.g. emergency VPN use, metadata removal from posts) and changes (increase) in real-time misinformation and digital warfare.

Findings show that there is an urgent need for more preparedness for specific crisis context communication and mechanism strategies.

Possible extension

- Evaluating the real world impact of shared security advice
- Investigate how misinformation spreads in conflict-related situations

Recommendations

Advice

- Focus on actionable and context-specific advice over generic security guidelines
- Advice tailored to crisis situation, especially in war zones
- Enhance digital literacy on misinformation
- Clearer indicators of trustworthy sources

Responsibility

- Responsibility is with the social media platforms → Design implications
- Develop better ways to verify and flag misinformation in real-time
- Content moderation policies
- Inclusion of governments: Crisis plans

Misinformation can be just as dangerous as more direct forms of cyber threats.

Areas of Inquiry

Within HCI

- Usable security & privacy design (e.g. accessibility and user experience)
- Information dissemination
- Crisis informatics (How people seek, verify, and act on information in times of crisis)
- User perception & decision-making (User protization of security recommendations)
  - HCI work on cognitive biases and behavior in security contexts

Outside of HCI

- Cybersecurity & intelligence (Digital warfare)
- Social / Political Science (Misinformation, trust, public response)
- Communication Studies
- Public Policy (Governance)

HCI Research Contributions

Extends HCI research by demonstrating how security advice functions in a high-stakes crisis context.

Combines practical user behavior and design challenges with different actors and levels (e.g. government, organizations).

Fills gap in existing research in that it extends research on user behavior to high-stakes and a crisis situation.

Main Takeaways

Misinformation as a security threat: In hectic crisis scenarios, users may be at even greater risk. Some security recommendations during the invasion were false, posing a risk to users.

Wide circulation of misinformation: Along with news on a rapidly evolving situation, false claims (e.g. hacking of Signal) were also spread quickly and widely.

Missing crisis-specific information: Most advice was generic. It was on the users to determine priority and relevance in a high-stake situation.

Conflicting information: Many government and international organizations shared in parts conflicting advice.

Crisis-specific advice ended up being the most relevant.

Discussion questions

What lessons can be learned from this case study about improving the dissemination of critical security and privacy information during future crises or conflicts?

How did the security and privacy advice shared during the Russian invasion of Ukraine differ from advice typically given in non-crisis situations? What unique challenges or threats emerged?

Synthesis / Overview

- Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks
    - Expands on traditional privacy frameworks to address AI-specific concerns
- Analyzing Security and Privacy Advice During the 2022 Russian Invasion of Ukraine on Twitter
    - Highlights the role of social media in crisis communication for security and privacy
- Privacy & Security Subcommittee Overall
    - Plays a crucial role in advancing research that makes privacy and security more human-centered and accessible to diverse user groups