Prevention of Interpersonal HARMS Through Threat Modelling

KIERON IVY TURK, Ruhr Universität Bochum, Germany and University of Cambridge, UK
ANNA TALAS, University of Cambridge, UK

Academic research has highlighted a wide range of technology-facilitated interpersonal harms and their impact. However, very few technologies incorporate *safety by design* into their development, leading to novel technologies that facilitate gender-based harms. We present the HARMS framework as a potential solution, facilitating risk identification and mitigation through integration into existing threat modelling practices.

CCS Concepts: • **Security and privacy → Social aspects of security and privacy**; **Software security engineering**.

Additional Key Words and Phrases: Harm Prevention, Threat Modelling, Safety by Design

ACM Reference Format:
Kieron Ivy Turk and Anna Talas. 2026. Prevention of Interpersonal HARMS Through Threat Modelling. In *Proceedings of Socio-technical Imaginaries for Responsible Design (CHI SIRD 2026).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Digital gender-based harms, such as technology-facilitated domestic abuse and online harassment, are widespread misuses of ubiquitous systems that affect a large proportion of the world population. These harms come in many forms that are well understood by academics, the support sector, and survivors of harm due to research, expertise, and first-hand experience. However, they are rarely considered as a potential issue by device manufacturers and application developers, leading to the continued facilitation of digital harms.

We can retroactively identify and work to mitigate threats to systems, but this often requires multiple iterative stages of further development to become effective. For example, item tracking devices such as AirTags were created to locate lost items, but quickly became used for stalking and intimate partner abuse. After widespread media coverage, tracker companies developed several initial interventions, but these had a variety of issues identified through academic research [1, 4, 5]. Following further improvements, the stalking prevention features are now universal and have improved implementations, including a standard for unwanted tracker detection [3]. However, ideally these attacks will be prevented before they become widespread instead of only mitigated after hundreds or thousands of users have been harmed.

The goal is for companies to be able to identify and mitigate threats without requiring extensive studies and media coverage to highlight issues. One potential approach is to adapt *threat modelling* from security by design practices by providing a framework that fits into existing analysis methodologies with the express purpose of identification of interpersonal harms.

Authors' Contact Information: Kieron Ivy Turk, kieron.turk@cl.cam.ac.uk, Ruhr Universität Bochum, Bochum, Germany and University of Cambridge, Cambridge, UK; Anna Talas, anna.talas@cl.cam.ac.uk, University of Cambridge, Cambridge, UK.

Manuscript submitted to ACM

2 Industry Threat Modelling

Threat modelling is a methodology for proactively identifying security risks to a system during the design phase. This security-by-design practice reduces risk by allowing for threat prevention mechanisms to be designed into a digital service or product, instead of only attempting to mitigate attacks when they occur and are reported. Threat modelling is widely used in industry by security teams and consultants in combination with vulnerability reporting programs to identify and counteract attacks.

The methodology generally consists of up to five steps: describing the system that needs to be defended; identifying the relevant "threat actors" who may attack the system; identifying the potential threats; prioritising the threats found; and designing mitigations for the threats. The process is often "ad-hoc", with users moving freely between different stages to ensure a more complete threat analysis.

To facilitate the identification of different theoretical attacks, it is common to use a threat modelling framework in the identification phase. These give insights into different threat types that should be considered for each component of the system, broadening the range of threats found during analysis and ensuring a more complete understanding of possible attacks. For example, the STRIDE model introduced by Microsoft [2] covers Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, and Escalation of privilege, and is widely used to identify technical attacks. An alternative framework which can be used in tandem is LINDDUN [7], which explores a range of privacy threats rather than technical attacks.

2.1 The HARMS Model

Introduced by Turk et al. [6], the Human HARMS threat modelling framework comprises five components which cover the largely unexplored threat of interpersonal harms. The components are Harassment, Access/infiltration, Restrictions, Manipulation and tampering, and Surveillance. The model is largely based on technology-facilitated abuse with coverage of wider interpersonal harm scenarios.

This framework is used in an identical manner to pre-existing threat modelling frameworks, which facilitates integration into established security practices. By performing a threat analysis with HARMS at the same time as using STRIDE or LINDDUN, technology can prevent or impede common forms of technology-facilitated gender-based harm with minimal additional effort from tech companies.

3 Future Directions

Security by design practices provide a solid foundation for future *safety* by design methods. Threat modelling is a widely-used and effective initial approach which can be easily extended through HARMS or similar frameworks to pre-emptively prevent technology-facilitated harms at an early stage.

Acknowledgments

We thank our colleagues at the Cambridge Cybercrime Centre for their feedback. The Human HARMS research was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grants EP/W032473/1 and EP/T517847/1 and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 949127).

References

[1] Alexander Heinrich, Niklas Bittner, and Matthias Hollick. 2022. AirGuard – Protecting Android Users From Stalking Attacks By Apple Find My Devices. https://arxiv.org/abs/2202.11813. doi:10.48550/ARXIV.2202.11813

[2] Loren Kohnfelder and Praerit Garg. 1999. *The threats to our products.* Technical Report. Microsoft Interface.

[3] Brent Ledvina, Zachary Eddinger, Ben Detwiler, and Siddika Parlak Polatkan. 2026. *Detecting Unwanted Location Trackers.* Internet-Draft draft-ledvina-apple-google-unwanted-trackers-02. Internet Engineering Task Force. https://datatracker.ietf.org/doc/draft-ledvina-apple-google-unwanted-trackers/02/ Work in Progress.

[4] Kieron Ivy Turk and Alice Hutchings. 2024. Stop Following Me! Evaluating the Effectiveness of Anti-Stalking Features of Personal Item Tracking Devices. In *The 2024 European Symposium on Usable Security (EuroUSEC 2024)*. 21. doi:10.1145/3688459.3688477

[5] Kieron Ivy Turk, Alice Hutchings, and Alastair R. Beresford. 2023. Can't Keep Them Away: The Failures of Anti-stalking Protocols in Personal Item Tracking Devices. In *Security Protocols XXVIII*, Frank Stajano, Vashek Matyáš, Bruce Christianson, and Jonathan Anderson (Eds.). Springer Nature Switzerland, Cham, 78–88.

[6] Kieron Ivy Turk, Anna Talas, and Alice Hutchings. 2025. Threat Me Right: A Human HARMS Threat Model for Technical Systems. In *Security Protocols XXIX*. Springer Nature Switzerland.

[7] Kim Wuyts, Laurens Sion, and Wouter Joosen. 2020. LINDDUN GO: A Lightweight Approach to Privacy Threat Modeling. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 302–309. doi:10.1109/EuroSPW51379.2020.00047

Received 12 February 2026; accepted 19 February 2026