Homefront Helper: Exploring the Self-Disclosure Risks of Military Family Members

ANIRBAN MUKHOPADHYAY, Virginia Tech, USA
JIN-HEE CHO, Virginia Tech, USA
KURT LUTHER, Virginia Tech, USA

Military family members carry an unusual privacy burden. Their everyday posts about school pickups, homecomings, moves, and deployments are ordinary on their own but revealing in combination, and the guidance they receive is infrequent, inconsistent, and written for service members rather than for spouses and parents. We describe Homefront Helper, a generative AI probe that reads a consenting participant's own Facebook history, detects self-disclosures against a military specific risk taxonomy, aggregates them through a Contextual Integrity pipeline, and returns a report that includes narrative attack scenarios assembled from the participant's own posts. We are deploying it in a qualitative study with 10 to 12 military family members to understand their OPSEC training and lived experience, to test where current detection models succeed and fail, and to learn what actually moves someone to act on a privacy warning. We end with a discussion on how the work aligns with the workshop themes of using privacy frameworks and responding to the quickly evolving technology landscape.

Additional Key Words and Phrases: privacy, contextual integrity, OPSEC, military families, generative AI

1 Motivation

In February 2015, five U.S. military spouses received messages telling them that the sender knew everything about them, their husbands, and their children [1]. The messages cited home addresses, family details, and photographs, all gathered from social media. The operation was first attributed to ISIS and later traced to a Russian GRU unit [2]. However, nothing about the attack required a breach.

That case is a useful anchor because the harm did not come from any single post. It came from the relationship between posts, where the posts together describe where a family lives, when the house is empty, and who is in it. Operational security (OPSEC) deals with this problem but the tools and training available to families have not caught up. Spouses report receiving guidance through occasional briefings that are often outdated and aimed at the service member rather than the household [5]. Automated work on military exposure has largely studied service members themselves [3]. Recent AI tools for risky self-disclosure are promising but tend to judge one piece of content at a time, and they are rarely evaluated with the people they are meant to protect [4].

This is also a problem specific to this military family population, and it should be accounted for in the design. The obvious advice of sharing less is isolating for military families who are geographically dislocated and dependent on online community for support during deployment. So, telling a spouse to stop posting about their lives might get ignored.

We are approaching this through a system and a study. We built Homefront Helper, a generative AI based technology probe that analyzes a participant's real posting history, and we are putting it in front of military family members to see what they recognize, what they dispute, and what they are willing to change.

Authors' Contact Information: Anirban Mukhopadhyay, Virginia Tech, Blacksburg, Virginia, USA, anirban@vt.edu; Jin-Hee Cho, Virginia Tech, Department of Computer Science, Virginia, USA; Kurt Luther, Virginia Tech, Department of Computer Science, Virginia, USA.

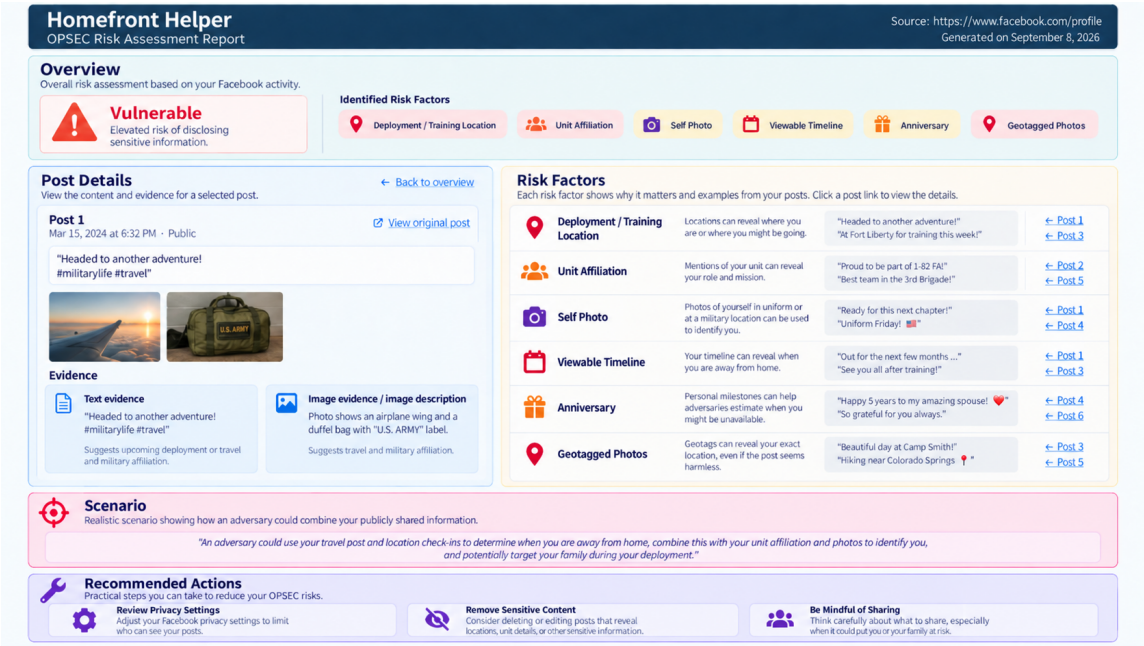


Fig. 1. Schematic of the Homefront Helper report presented to military family members. The Overview summarizes the profile-level assessment through an aggregate risk rating and the set of identified risk factors. The Risk Factors panel lists each detected disclosure category, explains why it may matter, and links to the posts that provide supporting evidence. Selecting a post opens the Post Details panel, which shows the original post content together with text- and image-based evidence used in the assessment. The report also presents an Attack Scenario illustrating how multiple disclosures could be combined by an adversary, followed by Recommended Actions suggesting practical steps for reducing the identified risks for some flagged posts.

2 Homefront Helper

Homefront Helper is a browser extension that can extract post details and generate a vulenrability assessment report. With consent, a study account is connected to the participant's Facebook network, the participant's historical posts and images are collected, and a report is generated and returned to them before the session. A schimatic diagram of the report is presented in Figure 1. We provide the system's reasoning along with the identified risk factors to provoke authentic reactions.

2.1 Risk Taxonomy

We start from an established scoring system for how much access a given piece of information gives an attacker, covering categories such as home address, geotagged photos, spouse name, contact phone, self photo, viewable timeline, professional skills, viewable network, rank, and current city [3]. We extend the original taxonomy with attributes that are specific to military life like unit affiliation, deployment and training dates, deployment and training location, and relocation timelines. These extensions carry high weights because they pose serious OPSEC risks.

2.2 Post Level Detection

We use the Kimi K2 open source, locally hosted, multimodal AI model to read each post's text, images, and metadata against that taxonomy. For every category it finds, it returns the source (text vs image contents of post), whether explicit, implied, image only, a short explanation, and ratings of military OPSEC relevance. Images matter as much as caption here as image elements often carry sensitive information.

2.3 Contextual Integrity Aggregation

The post level layer feeds a Contextual Integrity (CI) pipeline [6] instantiated for this setting. The sender is the profile owner. The subjects are the people in the posts, including the service member, the children, and the wider support network. The recipient is someone in the friend list but not very well known and the sender does not know a lot of their current status. The information type is not only declared personal data but inferred location, routines, deployment timing, etc. The transmission principle we test is whether a reasonable military family member would expect this bundle of facts to be inferable by a stranger from these posts.

The pipeline evaluates contextual cues, such as whether a post was meant for a close community rather than strangers, whether it exposes bystanders, whether it reveals a routine, and whether it becomes sensitive only in combination with earlier posts. It then makes a norm judgment about what becomes inferable, by whom, and at what confidence. For example, earlier school pickup posts and hometown mentions together, lets an outsider infer where the children are on weekday mornings.

2.4 Attack Scenarios

The report also contains short narratives that take several harmless looking details and combine them into a story about how someone could act on them. This part of the design is aimed at understanding if such scenarios make the disclosures more actionable. It might be effective in part because it is upsetting, and we treat participants' emotional reactions to it carefully.

3 Study

We are running a qualitative study with 10 to 12 military family members, including spouses, partners, and immediate family of active duty personnel and consent to sharing their post history. Sessions run about an hour. Participants first answer baseline questions before seeing anything the system produced, then open their own report and think aloud while moving through it, then reflect once the report is closed. We organize the work around the following three aspects.

Training and lived experience. What OPSEC awareness and training do participants actually have, how do they manage their sharing today, and what risks do they perceive? We collect this first so that the baseline is not shaped by the system's framing.

Model performance and user experience. Where do current detection models succeed and fail, particularly on multimodal reasoning, contextual awareness, and explanation quality? We ask participants to judge the evidence directly. Is the quoted snippet a fair read or a stretch. Is the image inference accurate. Did the category label mean what they assumed before they read the explanation. What did the system miss that they know is on their profile.

Awareness and mitigation. What makes someone willing to act, and how do they intend to mitigate? We provide some mitigation steps like changing the audience of posts from public to friends or omitting sensitive details from the text and blurring parts of the image. We ask directly which posts participants would refuse to change and why.

4 Alignment with the Workshop Themes

This work speaks most directly to the workshop's questions about frameworks and about keeping pace with technology. On the first, our contribution is an attempt to operationalize the CI framework. We have tried to turn CI into a pipeline with defined inputs and outputs, where the parameters are filled in by a model reading real posts, the LLM based norm judgment produces a disclosure level, and the resulting explanation is presented to the end users.

On the second theme, we think browser extension based probes are a practical answer to the pacing problem. Because the probe runs over the participant's own logged in view of a platform, it can be updated as model capabilities change and redeployed to the same people without rebuilding the study around a new dataset. That lets participants encounter new detection and explanation approaches as they mature, and lets us keep the same population in the loop across successive versions rather than sampling a fresh group each time capability shifts.

5 Statement of Interest

The first author recently completed his PhD in Computer Science at Virginia Tech working at the intersection of human AI collaboration, Open Source Intelligence (OSINT), and usable privacy and security. All the authors are interested in the challenges in privacy research and working on a future agenda. We would also like to bring some insights from our ongoing data collection for this project during the workshop and discuss ways to scale up such research.

6 Acknowledgment

We thank Dr. Debbie Bradbard for her help in reaching out to military family members and supporting participant recruitment. We also acknowledge funding support from the Commonwealth Cyber Initiative (CCI).

References

[1] 2018. 'ISIS hackers' threats against U.S. military wives actually came from Russian trolls. https://www.nbcnews.com/storyline/isis-terror/u-s-military-wives-threatened-russian-hackers-posing-isis-n872251

[2] 2021. Facebook says Iranian hackers targeted U.S. military personnel - CBS News. https://www.cbsnews.com/news/facebook-iran-hackers-us-military-defense-contractors-fake-accounts/

[3] Judson C. Dressler, Christopher Bronk, and Daniel S. Wallach. 2015. Exploiting military OpSec through open-source vulnerabilities. In *MILCOM 2015 - 2015 IEEE Military Communications Conference*. 450–458. doi:10.1109/MILCOM.2015.7357484

[4] Isadora Krsek, Anubha Kabra, Yao Dou, Tarek Naous, Laura A. Dabbish, Alan Ritter, Wei Xu, and Sauvik Das. 2025. Measuring, Modeling, and Helping People Account for Privacy Risks in Online Self-Disclosures with AI. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (May 2025), 1–31. doi:10.1145/3711029

[5] David W Lees. 2016. *Understanding effects of Operations Security (OPSEC) awareness levels of military spouses through the lenses of training and program management: A qualitative study.* Ph. D. Dissertation. Creighton University.

[6] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.