Technology and Power: Examining Imbalances Through Usable Security & Privacy Research

Kentrell Owens

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:
Franziska Roesner, Chair
Tadayoshi Kohno, Chair
Joseph Calandrino

Program Authorized to Offer Degree:
Paul G. Allen School of Computer Science & Engineering

University of Washington

Abstract

Technology and Power: Examining Imbalances Through Usable Security & Privacy Research

Kentrell Owens

Co-Chairs of the Supervisory Committee:
Franziska Roesner
Paul G. Allen School of Computer Science & Engineering

Tadayoshi Kohno
Paul G. Allen School of Computer Science & Engineering

Technology is often framed as a neutral tool, but in practice, it amplifies existing power imbalances, particularly when imposed by institutions rather than chosen by those most affected. This dissertation explores how surveillance, coercion, deceptive design, and informational asymmetry shape the experiences of individuals subject to technology they either did not choose or were unaware could harm them. Using methods from usable security and privacy research, I analyze case studies across the U.S. criminal legal system, the U.S. immigration system, deceptive design, and the modified Android app ecosystem to understand how design choices and data practices disproportionately harm those with the least power to resist them. These four case studies—which describe how technology exacerbates an existing power imbalance or generates a new one—fall within three categories of power imbalances: users and government entities, users and corporations, and users and other users. While these case studies showcase the impact of technology across differing contexts with difference consequences for users (e.g., incarceration, deportation, privacy harm, increased susceptibility to deception), they all highlight the unique ways that each technology affects power imbalances and how my work contributes to understanding them. This dissertation

ends with reflections on the broader implications of these imbalances, lessons learned, and potential strategies for mitigating harm in future technological development.

TABLE OF CONTENTS

Page

List of Tables iv

Chapter 1: Introduction 1
1.1 A motivating example 1
1.2 My motivation and approach 3
1.3 Users and government entities 4
1.4 Users and corporations 7
1.5 Users and other users 9
1.6 Technology and power imbalances 11

Chapter 2: Background and related work 12
2.1 Amplification theory 12
2.2 The role of technologists in social change 13
2.3 Related work 15

Chapter 3: Electronic monitoring smartphone apps 18
3.1 Introduction 18
3.2 Background on community supervision and electronic monitoring 20
3.3 Methods 22
3.4 Results 25
3.5 Privacy policy analysis 35
3.6 Putting our results in legal context 38
3.7 Discussion and conclusion 40
3.8 Technology as an amplifier 45

Chapter 4: Understanding experiences with immigration surveillance 47
4.1 Introduction 47

4.2 Background on surveillance of migrants 50
4.3 Methods 51
4.4 Results 55
4.5 Discussion and conclusion 73
4.6 Technology as an amplifier 77

Chapter 5: Deceptive design patterns in voice interfaces 78
5.1 Introduction 78
5.2 Background on deceptive design 81
5.3 Characterizing deceptive design patterns in voice interfaces 85
5.4 Surveying users' perceptions of deceptive voice patterns 93
5.5 Discussion & conclusion 105
5.6 Technology as an amplifier 108

Chapter 6: Modded WhatsApps 109
6.1 Introduction 110
6.2 Author contribution statement 112
6.3 Background on WhatsApp and modified Android apps 113
6.4 Interview study methods and findings 117
6.5 Contextualizing users' perceptions of mods 123
6.6 Discussion and conclusion 129
6.7 Technology & new power imbalances 135

Chapter 7: Conclusion and reflections 136
7.1 Comparisons of power imbalances and their impacts 136
7.2 Looking forward 139

Appendix A: Additional material for electronic monitoring smartphone apps 180
A.1 Review qualitative analysis codes 180
A.2 Privacy policy qualitative analysis codes 181

Appendix B: Additional material for understanding immigration surveillance 185
B.1 Interview protocol 185
B.2 Codebook 187

Appendix C: Additional material for deceptive design in voice interfaces 190
C.1 Survey instrument 190
C.2 Participant demographics 196

Appendix D: Additional material for modded WhatsApps 198
D.1 Interview study participant demographics 198
D.2 Testing procedure 199
D.3 Network traffic tracker analysis 203

LIST OF TABLES

Table Number Page

3.1 Permissions requested by each app. The leftmost permissions are label as "dangerous" by Android. 26

3.2 Third party trackers in monitoring apps (N=16) 29

5.1 Scenarios generated during our expert panel exercise 87

5.2 Heat map displaying the percentage of participants that chose a Likert item for each scenario. The scenarios that we intended as **deceptive** are bolded. . 96

5.3 CLMM model summary 104

6.1 WhatsApps used by participants. Participants could indicate multiple, and the three participants using the official WhatsApp were predominantly using WhatsApp mods. 121

6.2 Metadata of the apps we analyzed. The thumbprint is the SHA1 hash of the digital certificate used to sign the APK. The CrUX rank indicates that a website was in the Top X websites during December 2023 in Kenya. 125

6.3 Mod permissions that are not requested by the official WhatsApp. 128

6.4 Malware classificiation. AV-count is the no. of anti-virus engines classifying an app as malicious. 130

7.1 A summary of the chapters of this dissertation and their contributions 137

A.1 Summary of network traffic analysis. 183

A.2 EM app details, including Google Play Store installs and additional usage information. 184

C.1 Distribution of scenario responses. Participants (n=93) were randomly assigned three scenarios to evaluate. 196

C.2 Demographic information of the participants. Participants were able to select multiple levels for race/ethnicity and employment status. 197

D.1 Demographics of participants. 198

D.2 Bolded domains names are the ones in the ad/tracker list [183] from Ublock Origin. The "X" mark indicates the network traffic for the app in a given column contained the domain name in the row. 203

ACKNOWLEDGMENTS

It takes a village to raise a child, and I would not have been able to complete this dissertation without the support of multiple people.

I would like to thank my advisors Tadayoshi Kohno and Franziska Roesner. To paraphrase Langston Hughes (my favorite writer) in his poem 1922 "Mother to Son": "[the PhD process] for me ain't been no crystal stair" [175]. The PhD has been challenging in ways that I did not anticipate, and I have learned a lot about myself. I've also learned how to be a good researcher, mentor, and teacher, from my advisors. Due to their mentorship, I feel fully-equipped and motivated to pursue my current goal of becoming tenure-track computer science faculty, and I could not ask for much more than that! I would like to thank Joseph Calandrino and Martin Nisser for serving on my committee and offering great feedback and asking thought-provoking questions.

I would like to thank all of my research collaborators on work I did during my PhD, including William Agnew, Anita Alem, Olabode Anise, Adam Aviv, Abeba Birhane, Myra Cheng, Dave Choffnes, Camille Cobb, Lorrie Cranor, Yael Eiger, Pardis Emami-Naeini, Erin Freiburger, Johanna Gunawan, Kurt Hugenberg, Ryan Hutchings, Pratyusha Ria Kalluri, Amanda Krauss, Nino Migineishvili, Collins Munyendo, Basia Radka, Franziska Roesner, Mattea Sim, Luca Soldaini, Faith Strong, Blase Ur, and Shaoqi Wang. I've learned so much from you all!

I would like to thank all the former and current members of the UW Security and Privacy Research Lab that I have had the pleasure of meeting (and in some cases, befriending), including: Arka Bhattacharya, Maddie Burbage, Christine Chen, Inyoung Cheong, Kaiming Cheng, Camille Cobb, Aarushi Dubey, Yael Eiger, Pardis Emami-Naeini, Ivan Evtimov,

Michael Flanders, Chris Geeng, Gregor Haas, Rachel Hong, Umar Iqbal, Karl Koscher, Ada Lerner, Rachel McAmis, Alexandra Michael, Peter Ney, Mattea Sim, Lucy Simko, Anna K. Simpson, Miranda Wei, Henry Wong, Tina Yeung, and Eric Zeng. They made the lab and wonderful and welcoming place for me to grow as a researcher and a human. Thank you to David Kohlbrenner for the advice, honesty, and legal expertise he has offered over the years. Thank you to Nirvan Tyagi for his feedback and support on my postdoc job talk, the cookies that *he* baked (SIKE; thanks to Rachel for the cookies!), and his shoe recommendations.

Thank you to the CSE grad advising team (Elise Dorough, Joe Eckert, Christopher Nagle, Les Sessoms) for trying to make the Allen School as good of a place as it can be for grad students. We love y'all! Shoutout to UAW 4121, and all the amazing organizers that fight for our wages and healthcare!

I would like to thank my mentors that supported me before my PhD program, including Dr. Martin Carlisle (who hired me as a research assistant to work on picoCTF), Dr. Lorrie Cranor (who gave me my first opportunity to do usable security and privacy research), my undergraduate history professor, Dr. Alexander X. Byrd (who taught me so much about Blackness and how to love it), my high school English teacher, Dr. Nahla Mary Beier (who instilled within me a love of literature and taught me the importance of citation), and my undergraduate applied math instructor, Dr. Travis B. Thompson. I can confidently say I would not be graduating with a PhD in computer science now if it were not for Travis' enormous support and encouragement.

Thank you to Nite-Writers, a virtual co-working group that has helped me finish several projects in this dissertation. Thank you to the Black Graduate Student Association members (especially Kyle Johnson and Ngozi Ezeokeke, with whom I restarted the organization) and GSEE for helping me build community and feel like I'm not alone despite often being the only Black person in spaces I'm in. Thank you to Nicole Sullivan for crowdfunding a grill for me as a birthday gift!

I have to give a special acknowledgment to the following folks that have been there for me in big and small ways over the years and taught me the meaning of friendship and love. Some of us are frequently in touch, and some of us barely talk; regardless, everyone here had a powerful and positive impact on my life, including: Anita Alem, Aurelia Augusta, Leila Blair, Filip Drozdowski, Devan Evans, Brian Hou, Giselle Jhunjhnuwala, Sakiko Krishna, Alisa Liu, Ayush Narayan, Lucille Njoo, Maggie Oates, Nicola Park, Jonathan Price, Ewin Tang, Stephanie You, and a few people that I know would rather not be named here. A special shoutout to Nichole Williams, whose support has been invaluable during the past 4 years.

I would like to thank all my family members, particularly my mother, Sharon, who has supported my adventures over the years, even though they might not have always understood what exactly I was doing or what I was working towards. Lastly, I would like to thank the best friend I have ever had—the late, great Rishi Banda—for teaching me what I can do and encouraging me to go for it.

I received generous financial support from the ARCS Fellowship, the GEM Fellowship, and a UW College of Engineering Dean's Fellowship.

DEDICATION

To the least among us.

Chapter 1
INTRODUCTION

Reverend Dr. Martin Luther King, Jr. describes power as "the ability to achieve a purpose. Whether or not it is good or bad depends on the purpose" [41]. There are many existing power imbalances in society—some based on social factors like class, disability, gender, sexuality, race, or combinations of these [66]. Other power imbalances might be based on a legally-codified hierarchy, such as a judge having the authority to sentence someone [85]. While some scholars have argued that technology can create new power imbalances that did not previously exist [178, 191], others posit that technology typically amplifies existing social and political power imbalances [272, 297]. For example, technology might exacerbate an existing power imbalance between users and government entities (e.g., when there is government surveillance of individuals), resulting in more power for government entities and more severe risk and harms to users. I'll illustrate this scenario by providing an example.

1.1 A motivating example

Imagine you are at a protest exercising your constitutional right to assemble [7], and, suddenly, a police officer nearby is hit by some sort of projectile. You and several people near you are arrested and taken to jail. Within 24 hours you have your arraignment and are taken before a judge. You, knowing that you did not throw any projectile, plead not guilty. You tell the judge "I did not do this; I'm innocent." The judge (who may be sympathetic to your claim) says "Well, we'll review the body camera footage and discuss this all on your court date." The judge, who is an advocate for bail reform [270], decides that instead of making you pay a cash bail (or releasing you on your own recognizance) they will have you install an electronic monitoring app [222]. This app will track your location at all times and require

you to conduct remote check-ins with a supervisor. The check-ins will require you to do facial and voice verification using the app. If you fail to comply (as determined by the app) with the conditions (e.g., a home curfew) of your release, you may be arrested and brought back to jail. After presenting this information, the judge asks if you consent and requires you to sign a form indicating that you consent to this pretrial electronic monitoring. You have recently fallen on hard financial times and want to avoid paying bail, and although you have concerns about this app—particularly the many unknowns regarding the data it collects from your phone and who has access to that data—you sign the form and are released until your court date.

Smartphone apps, such as the one described in this example, are increasingly being used in the U.S. for electronic monitoring (EM) of people on probation, parole, pretrial release (as in the above example), or people in the juvenile or immigrant detention systems [45, 113, 120, 214, 221]. EM has typically been placed on people deemed "high risk," but carceral tech companies are marketing their apps as a low-cost and efficient way to expand the scope of surveillance to include "low risk" people as well [269].[1] People made to use EM apps often must pay regular fees to the app companies, do frequent biometric verifications, and ensure their devices do not run out of battery [221]. Failure to meet the conditions of one's release (as determined at least in part by the app) could lead to re-incarceration [113]. These EM apps exist within the broader context of community supervision [152], an existing power imbalance in the U.S. criminal legal system. I describe community supervision in more detail in Section 3.2.1. At a high level, EM apps exacerbate a power imbalance between users and government entities in ways that are disadvantageous for users coerced to use the apps.

[1] As one company put it: "This population of lower risk individuals or 'invisibles' is virtually out of sight in between check-ins with case managers and supervisors. While many lower risk individuals who are currently under this supervision do not require an ankle bracelet, there is community concern that a segment of this population should be under an additional veil of monitoring and supervision." [269]

1.2 My motivation and approach

In this dissertation I focus on multiple power imbalances, including ones between users & government entities (Chapter 3 & Chapter 4), users & corporations (Chapter 5), and users & other users (Chapter 6). In all the tech-impacted power imbalances I study, I focus on the risk and harms to the entity with less power—namely, users. This focus arises from a personal, ideological lens grounded in the idea that we should prioritize and focus on the needs of the most vulnerable among us. The narratives and concerns of those in power need no amplification; we know them, we constantly hear them, and they likely shape our own views of the world. Those with less power rarely have their needs and concerns elevated or centered; I hope to do this in my work.

Power imbalances that are exacerbated or created by technology often have security & privacy implications, such as facilitating surveillance or generating informational asymmetry. These imbalances can have severe impacts on users already facing marginalization or vulnerability in society. This is especially true when people are coerced to use (Chapter 3 & Chapter 4) or interact with (sometimes unknowingly) technology that they did not choose to use themselves (Chapter 6). Additionally, the modality of a certain technology may make users unable to recognize when they are being intentionally deceived (Chapter 5). My work aims to mitigate these power imbalances by increasing transparency, both (1) for users regarding how technology functions and (2) for developers and policymakers about technologies' impacts on users. Improving transparency for users enables them to make informed decisions regarding how they interact with a given technology (even if they are unable to opt-out). For developers and policymakers, understanding people's experiences using technologies can help them better design or better regulate technologies to reduce their harms to users. As a usable security & privacy researcher, I combine techniques from human-computer interaction research (e.g., qualitative methods) and security & privacy research (e.g., threat modeling, mobile app analysis). I use a variety of methods in this work, including qualitative analysis (of interview data, survey data, or online posts), Android app analysis (e.g.,

analyzing permissions, libraries, and network traffic), and legal analysis (via collaboration with a legal scholar).

This dissertation presents four cases studies of how technology makes existing power imbalances more severe or generates new ones. These case studies are instructive and highlight how technology can make already challenging circumstances much worse or how technology can be disruptive in undesirable ways. I will now introduce the four chapters of my dissertation by organizing them according to the type of power imbalances they address, summarizing their motivation, methods, and findings, and highlighting technology's impact on power for each chapter.

1.3 Users and government entities

Chapters 3 and 4 focus on tech-impacted power imbalances between users and government entities in the context of the U.S. criminal legal system and the U.S. immigration system. Government entities leverage technology (in these cases, primarily electronic monitoring apps) to surveil users in ways that were previously not possible—amplifying the power imbalance.

1.3.1 Electronic monitoring smartphone apps

Chapter 3 describes our work on electronic monitoring apps, from which the above example draws. A version of this work, coauthored with Anita Alem, Franziska Roesner, and Tadayoshi Kohno, was presented at the 31st USENIX Security Symposium in 2022 [222]. Despite the high-stakes nature of these apps, prior to our work we knew of no external audit evaluating their monitoring mechanisms, accuracy, or user impact. These apps have gained visibility due to prior reporting [113, 120, 221] but had not received noticeable attention from the computer science research community. To understand what type of privacy-related and other risks might be introduced to people who use these applications, we conducted a privacy-oriented analysis of 16 Android apps used for electronic monitoring. We analyzed the apps first technically, with static and (limited) dynamic analysis techniques. We also

analyzed user reviews in the Google Play Store to understand the experiences of the people using these apps and the apps' privacy policies. We found that apps contain numerous trackers, the permissions requested by them vary widely (with the most common one being location), and the reviews indicate that people find the apps invasive and frequently dysfunctional. We ended the paper by encouraging mobile app marketplaces to reconsider their role in the future of electronic monitoring apps, and computer security and privacy researchers to consider their potential role in auditing carceral technologies.

Takeaways about technology and power We analyzed EM apps, in part, to provide transparency regarding the apps' behavior and understand if any behavior we observed might violate legal limits. Through our collaboration with a legal scholar, we learned that when one signs the document given to them by the judge consenting to the conditions of their release under electronic monitoring (as described in the example scenario in Section 1.1), this essentially serves as a blanket consent. Given this interpretation, there appear to be no legal limits on the apps' behavior. The apps' third-party developers could—on their own or at the instruction of the agency that hired them—update their software to collect significant data that is irrelevant to the apps' functionality without consequence. Relative to ankle monitors, EM apps combined with this model of "consent" facilitate more invasive surveillance than previously possible and exacerbate the previously existing power imbalance. Moreover, the negative impact on users due to EM apps is more severe because EM apps add additional supervision conditions to the already burdensome system of community supervision—increasingly the ways and likelihood that people under EM will fail to comply with these conditions and be incarcerated.

1.3.2 Surveillance of migrants in the U.S.

In Chapter 4, I present our follow-up work on electronic monitoring apps by focusing on the app with the highest usage: `BI SmartLINK`. A version of this work—coauthored with Yael Eiger, Basia Radka, Tadayoshi Kohno, and Franziska Roesner—was accepted for publication

at the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT) [224].

People attempting to immigrate to the U.S. (through a port of entry or other means) may be required to accept various forms of surveillance technologies after interacting with immigration officials. In March 2025, around 160,000 people in the U.S. (and 9150 in the Seattle area [273]) were required to use a smartphone application—`BI SmartLINK`—that uses facial recognition, voice recognition, and location tracking; others were assigned an ankle monitor or a smartwatch. These compulsory surveillance technologies exist under Immigration and Custom Enforcement's (ICE) Alternatives to Detention (ATD) program, a combination of surveillance technologies, home visits, and in-person meetings with ICE officials and third-party "case specialists" [6]. For migrants in the U.S. who are already facing multiple other challenges, such as securing housing, work, or healthcare, the surveillance technologies administered under ATD introduce new challenges.

To understand the challenges facing migrants using `BI SmartLINK` under ATD, their questions about the app, and what role technologists might play (if any) in addressing these challenges, we conducted an interview study (n=9) with immigrant rights advocates between July 2024 and January 2025. These advocates have collectively supported thousands of migrants over their careers and witnessed firsthand their struggles with surveillance tech under ATD. Among other things, our findings highlight how surveillance tech exacerbates the power imbalance between migrants and ICE officials (or their proxies), how these technologies (negatively) impact migrants, and how migrants and their advocates struggle to understand how the technologies that surveil them function. Our findings regarding the harms experienced by migrants lead us to believe that `BI SmartLINK` should not be used, and these harms fundamentally cannot be addressed by improvements to the app's functionality or design. However, as this technology is currently deployed, we ended the paper by highlighting intervention opportunities for technologists to use our findings to make these high-stakes technologies less opaque for migrants and their advocates.

Takeaways about technology and power Migrating to the U.S. is already a challenging process. As of May 2025, refugee resettlement has been entirely stopped (with the exception of white South Africans [295]), and asylum requests at the southern U.S. border have been suspended [193]. `BI SmartLINK` exacerbates the challenges migrants face by subjecting them to increased surveillance and potential abuse at the hand of ICE officials (or their proxies) and by negatively impacting their mental health and ability to maintain a job, secure housing, and build community. We solicited questions from advocates about `BI SmartLINK` and learned that people are concerned about several things, including location tracking, surreptitious data collection, and risks to others nearby their phone. These questions lay the foundation for building an automated tool to 1) answer these questions, 2) either affirm or quell concerns about the app's data practices, and, hopefully, 3) mitigate this power imbalance through increased transparency.

1.4 Users and corporations

Corporations hold power over users due to their relative access to and control of resources. Chapter 5 describes how corporations can leverage their power over interfaces they control (particularly voice interfaces) to deceive users (e.g., by making it harder for users to choose settings that are better for their privacy).

1.4.1 Deceptive design patterns in voice interfaces

Chapter 5 describes our work on deceptive design patterns in voice interfaces. A version of this work, coauthored with Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner, was presented at the European Symposium on Usable Security in 2022 [225].

Deceptive design patterns (sometimes called "dark patterns" or "manipulative design patterns") are user interface design elements that may trick, deceive, or mislead users into behaviors that often benefit the party implementing the design over the end user. Prior work has taxonomized, investigated, and measured the prevalence of such patterns primarily in

visual user interfaces (e.g., on websites). However, as the ubiquity of voice assistants and other voice-assisted technologies increases, we must anticipate how deceptive designs will be (and indeed, are already) deployed in voice interactions. This paper made two contributions towards characterizing and surfacing deceptive design patterns in voice interfaces. First, we made a conceptual contribution, identifying key characteristics of voice interfaces that may enable deceptive design patterns, and surfacing existing and theoretical examples of such patterns. Second, we presented the findings from a scenario-based user survey with 93 participants, in which we investigated participants' perceptions of voice interfaces that we consider to be both deceptive and non-deceptive. While we found that participants considered scenarios we intended to be deceptive on average more problematic *relative* to the non-deceptive scenarios, we also found that overall, the majority of participants did not view our *intentionally* deceptive scenarios as problematic. This highlights the elevated potential for corporations to use voice interfaces as a form of power to deceive users without users even recognizing that they are being deceived.

Takeaways about technology and power Corporations control the choice archicture for visual interfaces; in other words, they dictate the controls users can access [139]. We show that voice interfaces, due to their unique properties, offer new pathways for expanding this existing power imbalance. Regulation on deceptive design is lacking regarding patterns in voice interfaces, and these deceptive design patterns are already in place (e.g., requiring different interfaces for subscribing and unsubscribing to premium services). As the results of our survey study highlight, people do not recognize some intentionally deceptive practices as deceptive, making them susceptible to deception in voice interfaces (perhaps moreso than in visual interfaces). We hope that regulators and corporations—that may be unintentionally using deceptive design patterns in voice interfaces—can use our findings to better protect users and mitigate this power imbalance.

1.5 Users and other users

Typically, when users interact with a specific social media or messaging platform through a smartphone app, they can have confidence that the capabilities of the app that they are using and those of the app that others are using are the same. For example, in the context of messaging apps (e.g., WhatsApp), read receipts are assumed to be symmetric. If I have read receipts enabled, then I can see other users' read receipts. If I do not have them enabled for my device, then I am unable to confirm if others have read my message. This is normally a valid assumption, but *modified* (also known as *modded*) apps complicate this. For example, individuals who use the official WhatsApp may (without their own knowledge or consent) interact with people who use a modified version in ways that violate their trust in the security & privacy promises of the official WhatsApp.

1.5.1 Risks and benefits of modified versions of WhatsApp

In Chapter 6 I describe work about how modified versions of WhatsApp create a new power imbalance by giving some individuals (i.e., those that use the modified WhatsApp clients) more information and control than others (those that use the official WhatsApp). A version of this work—coauthored with Collins W. Munyendo (who was a co-first author), Faith Strong, Shaoqi Wang, Adam J. Aviv, Tadayoshi Kohno, and Franziska Roesner—was presented at the 46th IEEE Symposium on Security and Privacy in May 2025 [210]. In this dissertation I describe our work on modified versions of WhatsApp, with an emphasis on the app analysis portion of the work (my primary contribution) rather than the user study.

WhatsApp is the most popular social messaging platform, and modified versions (or "mods") of the official WhatsApp are increasingly popular, particularly outside the West. Mods advertise themselves as having additional unique features relative to the official WhatsApp, such as the ability to send larger files, download images and videos from others' statuses, and block all incoming calls. However, some of these features, e.g., retaining deleted messages and statuses, enable mod users to subvert the privacy of others, and have the

potential for serious security and privacy harms.

In this study, we explored user expectations of WhatsApp mods through an interview study (n=20) of mod users in Kenya, one of the countries with the highest WhatsApp mod usage. To understand how users' expectations of WhatsApp mods align with the apps' behavior, I identified and analyzed 13 instances of the most common mod (GB WhatsApp). While WhatsApp mods contained the features that they claimed to offer, some participants in the interview study incorrectly believed that features currently available in the official app only existed in mods. Additionally, several mods were significantly over-permissioned compared to the official WhatsApp, despite participants believing that they requested the same permissions as the official app. While almost half of participants indicated they trust mods more than the official WhatsApp, we found that several mods were labeled by VirusTotal [25] as malicious, and two mods contained the "Triada Trojan," which is known to be distributed via WhatsApp mods [201]. Although WhatsApp mods have some features that users appreciate that do not have security & privacy risks for other users (e.g., blocking all incoming calls, an anti-spam/harassment feature), they pose risks to mod users (e.g., malware) and users of the official WhatsApp (e.g., informational asymmetry).

Takeaways about technology and power Our work shows that WhatsApp mods' features introduce a new power imbalance between WhatsApp users, creating an informational asymmetry. Users of the official WhatsApp may not even be aware of the existence of mods and may act in ways that put an ill-advised amount of trust in the user interface and security & privacy promises of WhatsApp. Users of the official WhatsApp may even be incentivized to use mods to reduce this power imbalance, despite the potential risks to them. This creates a negative feedback loop, in which more users want to use WhatsApp mods because they know other people are using WhatsApp mods, resulting in an overall increase in risks to more users. By studying the ecosystem of modified versions of WhatsApp, we hope to reduce this power imbalance by increasing awareness to WhatsApp and users about the risks and benefits (perceived or real) of using WhatsApp mods.

1.6 Technology and power imbalances

Collectively, the chapters in this dissertation elucidate how *usable security & privacy research methods are useful for critically evaluating how technology can create new power imbalances and make existing power imbalances much worse for those with less power.* This research approach is also helpful for identifying ways (e.g., through increased transparency) of mitigating these power imbalances. For the specific technologies covered in the four case studies included in this dissertation, my work explored if these technologies' promises outweigh their harms and risks to users, and, if they do, providing recommendations to mitigate (or eliminate) their harms and risks. In this dissertation, I show that technology can make a bad situation much worse (especially for vulnerable people), users may accept risks to themselves if they perceive that tech offers them a power advantage over other users, and, even in scenarios in which people are coerced to use technology, researchers can provide valuable insights that can improve the lives of those with less power—either through increased transparency for users, developers, or regulators.

The remainder of this dissertation is organized as follows. Chapter 2 provides relevant background about technology and power alongside prior work relevant to computer security & privacy for marginalized groups. Chapters 3 to 6 present individual research projects. In each of these chapters, the word "we" refers to work done by myself and my collaborators on each research project (referenced above). Chapter 7 concludes this dissertation and shares reflections and open questions. Appendices A to D includes additional materials, such as surveys, codebooks, and demographic tables, from each of the chapters.

Chapter 2
BACKGROUND AND RELATED WORK

In this section, I describe background information that is relevant for understanding my motivation and research methods.

2.1 Amplification theory

In Chapters 3 to 5, I present case studies in which technology amplified existing power imbalances. The idea that technology amplifies existing social forces is not a new one. In the 2002 publication "Real-time politics: The Internet and the political process" Philip E. Agre discussed the role that the Internet would have in politics; in particular, Agre challenged the techno-deterministic [298] notion that technology (the Internet in this case) "imprints its own logic on social relationships" [31]. Instead Agre presents an "amplification model" which posits that "the Internet changes nothing on its own, but it can amplify existing forces, and those amplified forces might change something" [31].

In the context of technology's impact on K-12 education, Warschauer et al. studied the impact of technology on educational outcomes in eight high and low socioeconomic status (SES) schools [282]. Introducing technology—namely, information and communication technologies (ICTs)—was presented as an approach to reduce the "digital divide." Warschauer et al. found no evidence that the introduction of ICTs reduced or mitigated disparate educational outcomes in the schools they studied. Moreover, they found that introducing ICTs served "to amplify existing forms of inequality" [282]. Schools that were well-resourced with strong teachers were able to use ICTs to improve student outcomes, but struggling schools were impacted by a variety of complex social factors causing the introduced ICTs to either have a zero or negative impact on outcomes. Although ICTs were initially emphasized as

a solution to the social problem (as is the case for the technology described in Chapters 3 and 4) of educational inequality, the authors state that this emphasis on technology may draw important attention and resources away from other more important interventions.

I was first introduced to the idea of technology as amplifier by Kentaro Toyama's paper "Technology as Amplifier in International Development" published at iConference 2011 [272]. In this work, Toyama presents the consequences of amplifier theory for "information and communication technology for development" (ICT4D) researchers. Historically, ICT4D as a research discipline is based on the beliefs that access to technology is a vehicle for improving quality of life, and developing regions in the world need research focused on improving their access to technology to enable "development" [194]. Toyama rebuts these beliefs, stating that "technology tends to amplify existing inequalities." Toyama also points out that "technology cannot substitute for missing institutional capacity and human intent" and ICT4D work is most successful when it supports already successful development efforts [272].

The power imbalance between users and government entities is the focus of Chapters 3 and 4, and the defining aspect of this imbalance is surveillance. When discussing Simone Browne's work on the surveillance of Blackness [59], Ruha Benjamin notes that while "challenging a technologically deterministic approach, [Browne] argues that, instead of 'seeing surveillance as something inaugurated by new technologies'," we should consider it ongoing and supported by existing power imbalances (i.e., anti-Blackness) [50].

2.2 The role of technologists in social change

Power imbalances may be intrinsic to the way certain systems are designed (e.g., users and government entities); they may also be reflective of social problems (e.g., white supremacy). Technologists, like everyone, have a role to play in solving social problems. This section outlines different approaches technologists could take, while providing caveats about how they approach this work.

In their 2020 FAccT paper "Roles for Computing in Social Change", Abebe et al. suggest that computing research has an important role to play in addressing social issues [27]. They

categorize these roles into four different categories: computing as a diagnostic, computing as a formalizer (i.e., developing models for social problems with clear inputs and outputs), computing as a rebuttal (i.e., highlighting the limits of technology), and computing as a synecdoche. As a diagnostic, computing research can help us measure social problems and diagnose how they manifest in society. An example of this work is Joy Buolamwini and Timnit Gebru's seminal "Gender Shades" paper that diagnosed how facial analysis systems performed most poorly on women with darker skin [61]. In Chapters 3 and 6, we measure the technical behavior of apps to understand the risks they pose to users, and in Chapter 5 we diagnose how voice interfaces can be designed to deceive users. Computing as a synecdoche can reframe or clarifying existing social problems in a different light or to a different community of researchers. Virginia Eubanks' book "Automated Inequality" studied poverty through the lense of algorithms used to administer poverty policy [108], bringing attention to the issue of poverty while engaging computer scientists in discussions about it. Our work in Chapters 3 and 4 was among the first (if not the first) research publications presenting research on electronic monitoring apps and surveillance of migrants to computer science researchers. Studying carceral technologies from the perspective of technologists offers a new perspective that can (justly or unjustly) add weight to the issue and draw the attention of others that may have previously ignored it. This work (particularly the work in Chapter 3) has already begun to reach more technologists and policymakers; it has been covered by the Electronic Frontier Foundation [132] and New Scientist [155], and was also cited in a series of letters several prominent U.S. Senators wrote to electronic monitoring companies expressing concerns about the industry's abusive practices [103].

Work by Chelsea Barbaras highlights the potential for technologists to play an active role in resisting power imbalances exacerbated by technology through "refusal as resistance" [46]. Due to the high demand for technical skills and the privileging of technical knowledge over other ways of knowing, technologists are in "a powerful position to negotiate and challenge the underlying theories of change associated with a given data project" [46]. Barbaras outlines 3 common missteps that technologists take when investigating carceral technologies:

"(1) 'proving' harm, (2) adopting deficiency narratives and (3) optimizing harmful systems." The consequences of "optimizing harmful systems" also arise in Ben Green's work [131], which discusses how technologists who intend to do social good often fall short, in part because they often (wrongly) assume that technology-centric gradual reform is the way to achieve social good. Green also points out how technologists often have an abstract idea of "social good" but do not interrogate if what they are doing is truly "good" and for whom it might be good [131]. Outside the context of technology, there is high variance on what is even considered a social issue. While many people agree that white supremacy is an unjust power imbalance that should not be exacerbated by technology, a 2022 survey found that a majority of Republicans in the U.S. believe in "the great replacement narrative," which "provides the central framework for the global white supremacist movement" [67]. Technologists are not uniquely well-positioned arbiters of what "social good" means [123] and must be thoughtful and reflexive when conducting research to avoid missteps that could result in more harm than good. In our work on immigration surveillance in Chapter 4, we describe our approach to ensuring that we avoid these common pitfalls in Section 4.5.3.

2.3 Related work

2.3.1 Android app analysis

There has been prior security and privacy work focusing on analyzing Android applications to ensure that the work complies with regulations (e.g., COPPA, HIPAA, GDPR). Researchers have previously conducted similar privacy-focused analyses of Android apps. Feal et al. [112] conducted a privacy study of 46 parental control apps (aka "parentware") using similar static, dynamic, and privacy policy analysis techniques as in our work. They also provided legal context for their research (COPPA). They used a customized version of the Lumen Privacy Monitor [277], a tool that we also use in our app analysis in Chapter 3. They found that 11% of the apps transmitted personal data insecurely, 34% collected and shared personal information without the appropriate consent, and 72% shared data with third-parties without

mentioning their presence in their privacy policies. Overall they found that the apps lacked transparency and did not comply with regulatory requirements, even apps recommended by government-affiliated entities. Han et al. [137] conducted a similar analysis comparing the privacy of pairs of free and paid versions of consumer apps. They found that despite popular belief otherwise, (i.e., that paid versions of apps should have less tracking since there should be fewer ads), paid and free versions of apps had similar collection and sharing practices regarding sensitive data. Nguyen et al. [215] did a large-scale measurement study to evaluate if apps violate GDPR's explicit consent requirement; they found that many apps send sensitive data before consent was explicitly given. The dynamic analysis in Chapter 3 also revealed data being sent before consent was given (e.g., to Facebook).

2.3.2 Computer security and privacy for marginalized populations

Recent work in the computer security and privacy community has examined the unique security and privacy needs of several specific (and sometimes marginalized or vulnerable) populations, including refugees [258], undocumented immigrants [134], users outside of W.E.I.R.D. [143, 180] contexts [208, 209, 247], survivors of human trafficking [70] and intimate partner violence [69, 117, 141], older adults [118], teenagers [86], people with visual impairments [33], people with a low socioeconomic status [188, 235, 257], transgender people [177], and sex workers [199]. My work in Chapters 3, 4, and 6 contributes to this space by exploring the needs of groups (people under electronic monitoring, migrants, and Kenyan users, respectively) that have recently noticeably less attention.

In an interview study closely-related to my work in Chapter 6, Naveed et al. explored the privacy behaviors of 40 low-literate and low-income users in Pakistan [212]. The researchers found that some of their participants (17.4% of women and 35.3% of men) used modded apps (GB and FM WhatsApp), and some described leveraging the modded app features to preserve their privacy. For example, one participant reported freezing her "last seen" time so that her brother would not come in and scold her for being online at night. Among their participants, most people learned about the modded apps from friends or co-workers.

Related to the immigration surveillance work in Chapter 4, Guberek et al. conducted an interview study with undocumented people in the U.S. and asked them about their technology use, risk perceptions, and protective strategies [134]. While none of the participants described being monitored under Immigration and Customs Enforcement's (ICE) Intensive Supervision Appearance Program (ISAP) (which had not yet been launched when the interviews were conducted in 2017), they described a general fear of surveillance and a perception that ICE was constantly monitoring their online activity, mirroring work study the security & privacy needs of refugees [134]. Additionally, Austin Kocher wrote about how another app (`CBP One`) was imposed on migrants applying for asylum at the U.S.-Mexico border [170]. `CBP One` digitizes several forms required for entering the U.S. at a port of entry, tracks the location of migrants' phone, and uses facial recognition for identity verification. Kocher argued that while this app is marketed as streamlining the administrative process of applying for asylum, it actually introduces digital barriers for asylum seekers, both in its proper function and as a result of "glitches." [1]

[1]Shortly after President Donald Trump was sworn into office on January 20, 2025, it was announced that `CBP One` would no longer be able to schedule appointments, and all existing future appointments were cancelled [35].

Chapter 3

ELECTRONIC MONITORING SMARTPHONE APPS

Section 1.1 paints a picture of how someone might come to placed on an electronic monitoring app and how the app exacerbates an existing power imbalance (i.e., community supervision) between users and government entities in the U.S. criminal legal system. In this chapter, we'll explore electronic monitoring apps in more detail, along with their impacts on people coerced to use them, and the legal context in which they exist. The chapter is based on a paper coauthored with Anita Alem, Franziska Roesner, and Tadayoshi Kohno, that was presented at the 31st USENIX Security Symposium in 2022 [222].

3.1 Introduction

Smartphone apps are increasingly being used in the U.S. for electronic monitoring (EM) of people on probation, parole, pretrial release, or people in the juvenile or immigrant detention systems [45, 113, 120, 214, 221]. EM has typically been administered to people deemed "high risk," but prison industry companies are marketing their apps as a low-cost and efficient way to expand the scope of surveillance to include "low risk" people as well [269]. People made to use EM apps often must pay regular fees to the app companies, do frequent biometric verifications, and ensure their devices do not run out of battery [221]. Failure to meet the conditions of one's release (as determined at least in part by the app) could lead to re-incarceration [113]. Yet, despite the high-stakes nature of these apps, we know of no external audit evaluating their monitoring mechanisms, accuracy, or user impact. These apps have gained visibility due to prior reporting [113, 120, 221] but have received no noticeable attention from the computer science research community.

We conducted a privacy-focused analysis of a subset of smartphone EM Android apps

from a technical, human, and legal point of view. We identified 16 apps that are used by tens of thousands of people in the U.S.

We seek to answer the following research questions through our exploratory analysis of these apps:

- **What are the privacy-related technical properties of the apps, including what permissions they request and what network endpoints they contact?**

We analyze this question through static and (limited) dynamic analysis. We found that all apps but one requested fine-grained location access, and that the difference in the number of permissions requested by the most privileged app (14) and the least privileged app (0) was significant. Regarding network traffic, passive observers (e.g., ISPs) may be able to identify that someone is using an EM app based on the domains contacted.

- **What are the experiences and concerns of people using these apps?**

We investigate this question through a qualitative analysis of user reviews in the Google Play Store. We found that app reviews surface concerns about malfunctions, these apps' disruptiveness, and dissatisfaction with the *proper function* of these apps. Malfunctions discussed were mainly related to an inability to use the app to successfully perform a check-in—an important requirement of community supervision. Disruptions caused by the apps included 1) loud alerts in inappropriate settings (e.g., work or church) or at inappropriate times (e.g., they were asleep), 2) taking up significant resources on their smartphones, such as space and battery, and 3) causing the entire smartphone to crash or freeze, potentially jeopardizing an EM condition that their phone is always running and available.

- **What is the relationship between what is stated in the apps' privacy policies and the potential risks and harms surfaced by our first two research questions?**

We investigate this question through an analysis of the privacy policies. Three apps do not have a privacy policy available in the Google Play Store, indicating that they may be in violation of its user data policies [142]. Only 9 of 16 apps had a privacy policy that explicitly addresses the apps' usage, and we discovered that one app company may have taken down its privacy policy in response to public scrutiny. While the level of details regarding data collection vary, almost all the apps said that they share data with third-parties, sometimes for marketing or advertising purposes.

Given the answers to our research questions, we present a case study of the least and most privileged apps and discuss the legal landscape related to these applications, in partnership with a legal collaborator. Collectively, our work contributes the *first systematic analysis of the electronic monitoring apps ecosystem*, and we conduct this analysis from a technical, human and legal perspective. We provide recommendations for mobile app marketplaces to increase transparency, and to the computer security & privacy community to reduce the potential harms of carceral technologies. The following sections present relevant background information, our methods, and major findings in more detail.

3.2 Background on community supervision and electronic monitoring

3.2.1 Community supervision in the U.S.

The U.S. is the most incarcerated country in the world by both incarceration rate and total number of people incarcerated [249]. In 2020 there were 2.3 million people incarcerated. Recent polls indicate most adults in the U.S. believe that the prison and jail population should be reduced [22, 28]. However, they may be unaware of the related problem of community supervision. In the words of one district attorney, "mass supervision is the evil twin of mass incarceration" [56].

In 2020, approximately 4.5 million people in the U.S. were under "community supervision," which can include people on probation, parole, pretrial release, or people in the juvenile or immigrant detention systems [45, 152, 204]. People in these programs must comply with

conditions (typically 18-29 rules [80]) that could result in incarceration if violated. These conditions include things like passing regular drug tests (even if someone's conviction was not drug-related), curfews, paying a supervision fee, and complying with geofencing [152]. Because these rules are extensive, and difficult to follow, people often fail to meet them and return to prison or jail due to "technical violations"—things would not be considered a crime if the person were not under community supervision (e.g., failing to pay a fine, missing an appointment) [48]. Around one-fourth of admissions to state prisons in the U.S. are due to technical violations [250], and over half of the people incarcerated in the U.S. are in state prisons [249].

3.2.2 Electronic monitoring

Many people under community supervision are also under "electronic monitoring" or EM, also known as "e-carceration" [165]. While there is no national count of the number of people on EM, a Pew report [145] stated there were 131,000 people on EM in 2015, up 140% from 2005; that number is an under-count, as it only includes GPS and radio-frequency (RF) units. EM agreements may involve twice as many conditions compared to ones that do not involve an electronic monitor [284, 285].

Historically, EM has taken the form of an ankle monitor (GPS or RF-enabled), but smartphone apps are increasingly being used for EM. People made to use these apps have reported problems such as poor connectivity, general malfunctions, and false positive alerts sent to their EM supervisor (e.g., a probation officer) [113, 120, 221]. EM apps typically track location and are used to perform check-ins with EM supervisors, in addition to or in lieu of in-person meetings [113, 120, 214]. Check-ins might require people to face their phone's camera (for live facial recognition) or capture a photo or video of themselves. Check-ins might also use voice recognition and require people to read off a random string of numbers while facing their phones [221]. People using these apps may receive loud notifications, sometimes at random times, alerting them to complete a check-in. Similarly, apps may send loud, warning notifications caused by incorrect sensor data (e.g., location); one report indicated that these

have occurred while people are sleeping [221].

Smartphone apps can be used to impose stricter conditions of supervision than would be possible under physical surveillance. For example, in a civil parental rights case, a father was ordered by a juvenile court to submit to random smartphone breathalyzer tests five times per day using `Outreach Smartphone Monitoring`, one of the apps we analyzed; any non-compliance or failure to submit within 30 minutes of an alert was assumed to constitute a positive alcohol screen [20]. The appeals court found the father's failure to complete 993 tests, out of a total of 2,317 check-ins in the span of about one year, to support terminating his parental rights. Mandating a check-in *five times* a day is only possible because of the smartphone app (and its companion Bluetooth breathalyzer); such a condition would be virtually impossible if it required travel to a physical location. Unlike most apps, which are subject to an open market, these apps involve people being more or less forced to use them. That is, the apps are not being built for the people using them, but for the carceral system. As these apps continue to grow in usage and cause problems for the people coerced to use them [113, 120, 221], there is a pressing need for external auditing and accountability. Although there has been some reporting on this ecosystem, there has been no systematic analysis—we aim to close that gap.

3.3 Methods

We conducted static and limited dynamic analysis of 16 EM apps. Static analysis reveals what an app could *potentially* do by examining the app's code, and dynamic analysis reveals what an app *does* in controlled execution environments. We also qualitatively analyzed the apps' reviews in the Google Play Store and their privacy policies. We identified the 16 apps we analyzed from news articles, search engine results, and suggested similar apps in Google Play. We searched for combinations of terms like "smartphone apps," "electronic monitoring," "probation," and "parole" on different search engines. These apps were downloaded in or before August 2021.

Static Analysis To conduct static analysis, we downloaded the apps onto a device and extracted them via Android Debug Bridge. Examining the permissions an app requests, the third-party libraries it uses, and its source code (although obfuscated) can reveal information about the app's data collection practices, who could gain access to the data, and how they might use the data. We analyzed the output of MobSF (Mobile Security Framework), a mobile application static and dynamic analysis tool [87]. Among other things, MobSF presents an app's third-party libraries, decompiled source code, and geo-location (based on server IP address) for any domains detected in the code.

Limited Dynamic Analysis Our ability to dynamically analyze the applications under normal operating conditions is, unfortunately, limited, because we either cannot directly create accounts ourselves (n=14) or, in some cases, choose not to do so to avoid agreeing to any terms of service (n=2). In either case, we cannot test the apps as they are used in interaction with EM supervisors. This limitation of our investigation emphasizes again the limited transparency and accountability in this ecosystem.

Nevertheless, we conduct a limited dynamic analysis of pre-login application behaviors. While running each app, we accepted any requested permissions and interacted with the app until we reached a login screen, leaving and returning to the app several times.

To gain visibility into the content and security of the network traffic, we collected network traffic while using the app and conducted a machine-in-the-middle (MITM) attack (when possible) for decryption. Using a Nexus 5X device running Android 8.1 (API 27) with mitmproxy [83], Wireshark [116], and Lumen Privacy Monitor [234, 277], we installed each app on the device and ran it for 10 minutes while capturing network traffic in Seattle, Washington. We collected traffic twice for each app with both a rooted [151] and an unrooted device; some of the apps detected that the device was rooted (and displayed a notification accordingly), and we wanted to know if that detection impacted what network traffic was sent. We instrumented the device with our own root certificate (via mitmproxy) by adding the certificate to its system store. Using Wireshark allowed us to capture network traffic

that used protocols aside from the HTTP/S capture supported by mitmproxy; we also used it to verify that our network captures were working properly. Lumen's tracking of DNS transactions allowed us to attribute encrypted network traffic to specific apps, compensating for our lack of visibility into encrypted HTTP headers. After running each app, we deleted it from the device before installing the next one, verifying that it did not modify the phone's state.

App Review Qualitative Analysis For the user review analysis, we collected all 257 reviews available in the Google Play Store and conducted qualitative content analysis [227]. Two researchers independently read through all of the reviews, each making a broad list of topics people raised. They discussed the list and jointly created a code book matching topics to closely related themes. They iterated on this code book and reached consensus on the codes to use. Using these codes, one researcher coded all of the reviews and discussed ambiguous reviews with other researchers when necessary. Our goal in analyzing app reviews, as with other qualitative work, was not to draw generalizable conclusions about the prevalence of certain issues, but rather to identify and surface the set of issues that people encounter and write reviews about. Consequently we do not attempt to use the review data to make generalizable or statistical claims.

Ethical Considerations We applied for IRB approval through our institution and received official notification from the IRB that our work does not qualify as human subjects research. Nevertheless, to evaluate the ethics of analyzing public app reviews without author consent, we considered the guidelines created by Buck et al. [60] for ethical treatment of data from online sources. This study focuses on analyzing people's concerns with using these applications and studies discourse rather than the people themselves. Moreover, this collection of reviews does not appear to violate the Google Play Terms of Service [126].

We considered seeking people under EM who use these apps, and asking them if we could experimentally evaluate the properties of their apps while they used them. One of the

reasons we chose not to do this is that we considered it too difficult to ethically experiment with the apps of people currently under EM; this could introduce risks to them and cause friction with their EM supervisor.

We found that seven apps in our study (`Sprokit`, `Corrisoft AIR Check-In`, `Community Supervision`, `aCheck`, `BI SmartLINK`, `Omnilink FocalPoint`, `Telmate Guardian`) appeared, at the time of our research, to be in violation of the Google Play Store's user data policies [142]. Three of the apps (`Sprokit`, `Corrisoft AIR Check-In`, `Community Supervision`) requested access to sensitive permissions but did not have privacy policies linked on their respective Google Play pages; four of them (`aCheck`, `BI SmartLINK`, `Omnilink FocalPoint`, `Telmate Guardian`) had links to privacy policies, but the policies did not mention the smartphones applications. We notified the companies (aside from `Sprokit`, which we discovered was no longer available in the Google Play Store as of February 2022) with a deadline by which they must add an adequate privacy policy to their Google Play pages. Our plan was the following: if the changes are not made by the deadline, we would contact trusted contacts at Google who specialize in vulnerable populations for guidance on next steps. In late March 2022, we reached out to our contact at Google notifying them about five apps (those listed above aside from `BI SmartLINK` and `Sprokit`) that may be in violation of their policies. In Section 3.5 we describe the response from the developers of `BI SmartLINK`.

3.4 Results

We analyzed these apps' permissions, network traffic, and third-party library usage.

In the following sections, we present general findings for all apps before presenting case studies of the apps we determined to be the most privileged (regarding the data they can access) and least privileged.

Information Sources: Permissions Permissions determine the types of data apps can collect. To understand the privacy risks to people using these apps, we must first understand what types of data can be collected about them. People under EM are required to accept at

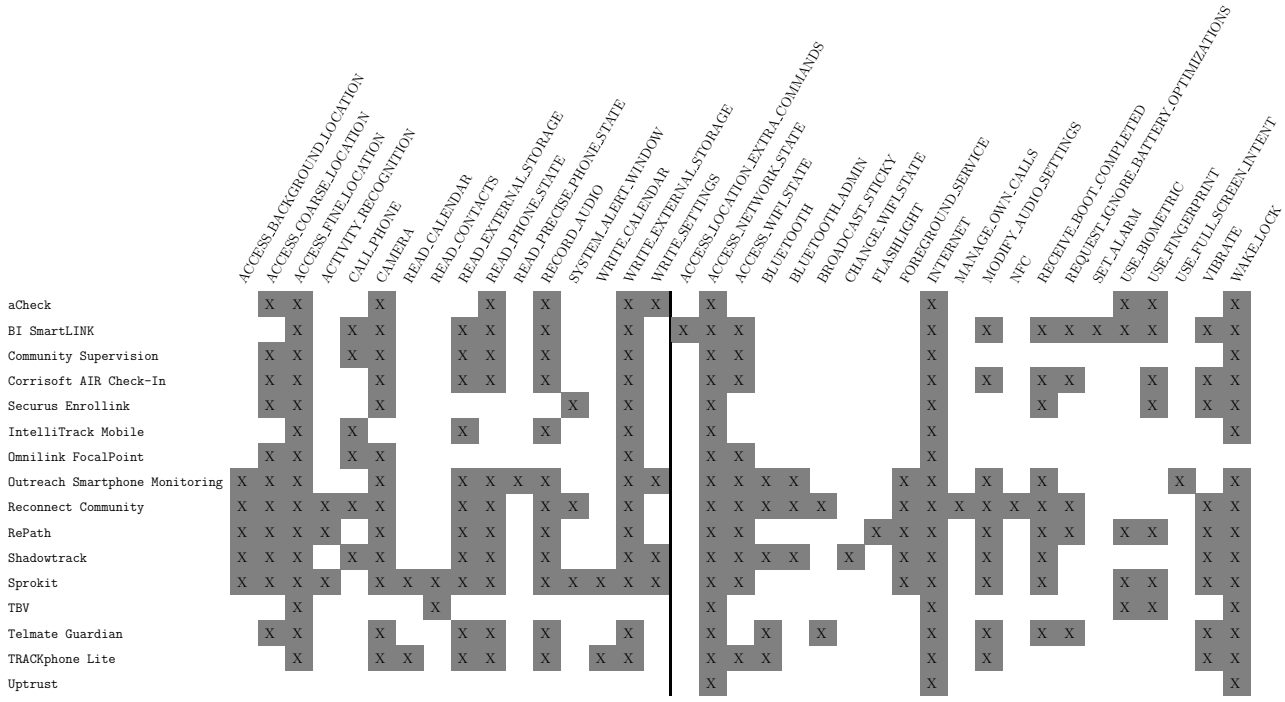

Table 3.1: Permissions requested by each app. The leftmost permissions are label as "dangerous" by Android.

least some, if not all, of the permissions requested by these apps. For example, some apps request permissions to offer certain features and continue to function if certain permissions are denied; other apps (such as `Telmate Guardian`) do not allow actions within the app (such as login) until all permissions have been granted.

Smartphone operating system permissions protect access to restricted data and restricted actions. Apps that request more permissions can send more data to supervisors and third-parties. By analyzing the distribution of permissions requested by these apps, we can compare them to the least-privileged app among them. If the least-privileged app has the same (or similar) goals as the other apps, it stands to reason that it may be able to serve as a standard for the "minimum number of permissions" necessary for other apps to function.

What is the prevalence of different *dangerous* permissions in these apps? Table 6.3 shows the Android permissions requested by the apps. The permissions on the left of the vertical line are labeled as "dangerous" in Android API documentation, while the rest in the table are considered "normal" permissions [91]. Dangerous permissions allow apps to access otherwise restricted data and take otherwise restricted actions. For the purpose of brevity and because they are not relevant to our investigation, we have excluded certain custom permissions, and phone manufacturer specific permissions.

The most common permission gave apps access to precise location information All EM apps but one (`Uptrust`) requested ACCESS_FINE_LOCATION, reflecting the centrality of location tracking to these apps' functionality. This permission enables apps to receive a location that is "as accurate as possible, ... sometimes as accurate as within 10 feet (a few meters) or better" [92]. It is notable that beginning with Android 10 (API Level 29, released September 2019), apps that wish to request a phone's location running in the background must request ACCESS_BACKGROUND_LOCATION [92]; only 5/16 monitoring apps did. However, as only 8.2% of Android devices use Android 10 [90], this may not affect these apps' ability to track background location on most devices. Most apps requested CAMERA (13/16) and RECORD_AUDIO (12/16), indicating potential use for biometric face or voice authentication or the use of video and audio data for other purposes.

A few apps requested permissions that did not have widespread use There were several dangerous permissions that only a few apps requested, indicating that they may not be necessary to offer capabilities similar to those offered by other monitoring apps. For example, only `Outreach Smartphone Monitoring` requests the READ_PRECISE_PHONE_STATE permission, which allows the detailed reading of information about phone state [91]. `TBV` and `Sprokit` are the only monitoring apps that requested the READ_CONTACTS, which—when combined with READ_PHONE_STATE—could allow supervisors to monitor whom someone talks to and how frequently they speak. Similarly, three of the apps request the

ACTIVITY_RECOGNITION permission, which reports if someone is in a vehicle, on a bicycle, running, or still [89].

Information Sinks: Third-Party Libraries While the presence of permissions reveals what type of data may be collected, the presence of third-party libraries reveals to whom collected data may be sent. Third-party libraries may have access to sensitive data about people using EM apps and may even monetize their use of the app. Smartphone apps typically include third-party libraries, which are sometimes referred to as SDKs (software development kits), although SDKs are broader in scope and often contain more than one library [101].

While two apps had no trackers, nearly all of the apps contained one or more Google analytics libraries MobSF [87], a mobile application analysis tool, uses Exodus [231], a tool to identify trackers in Android applications based on a list of known trackers. According to Exodus, a tracker is software that is meant to collect data about the person using a device or how the device is used; third-party libraries fall under this definition. Table 3.2 displays the trackers we found in the apps. Only two of the monitoring apps (`Omnilink FocalPoint` and `Corrisoft AIR Check-In`) contained no trackers at all. All of the remaining apps but `Shadowtrack` contained at least one Google-based analytics tracker.

Two apps use ad libraries, indicating the companies behind the apps might profit from the compulsory use of these apps `Telmate Guardian` contained the Flurry library, but appeared to only use its analytics capabilities and does not implement the code necessary to serve ads in the app. `Sprokit` appeared to contain the code necessary for Google AdMob and Facebook Ads SDKs to serve ads and monetize use of their app.

Two apps use Facebook Analytics and Login SDKs `Sprokit` and `Uptrust` both use Facebook Analytics and Login SDKs. This means that if someone logs into Facebook in the app, at a minimum the app gets access to their public profile and email address [111].

Tracker	**Type**	# **apps**
Google Firebase Analytics	analytics, databases, messaging, crash reporting	12
Google CrashLytics	crash reporting	6
Google Analytics	analytics	2
Microsoft Visual Studio App Center Analytics	analytics, push notifications	2
Microsoft Visual Studio App Center Crashes	crash reporting	2
Facebook Analytics	analytics	2
Facebook Login	login	2
Google AdMob	advertising	1
Facebook Ads	advertising	1
Facebook Share	content sharing	1
Amplitude	analytics, profiling	1
Segment	analytics, profiling	1
OneSignal	push notifications, messaging	1
Branch	analytics	1
Flurry	analytics, advertising	1
New Relic	analytics	1
UrbanAirship	analytics	1

Table 3.2: Third party trackers in monitoring apps (N=16)

Additionally, Facebook learns that this person is using the EM app.

Information Flows: Limited Dynamic Analysis Unlike permission and libraries, apps' network traffic reveals what they actually do. To understand the data sharing practices of

EM apps, we collected data while interacting with the apps. Specifically, we wanted to know what types of data are sent from the apps, to whom those data are sent, and if those data are sent securely. Understanding what types of data are sent allows us to examine the potential risks associated with those data and what data (if any) cross expected or legal bounds. while we would prefer to do a thorough dynamic analysis (e.g., collecting network traffic while using a tool to simulate user interaction with the app), our inability to create accounts and the lack of transparency in this ecosystem prevented us from doing so. Although the dynamic analysis was limited, it places a lower bound on what network traffic apps send.

Apps send traffic to their servers and some of the third-party services detected in their code Four apps (`Securus Enrollink`, `Omnilink FocalPoint`, `Corrisoft AIR Check-In`, and `TBV`) had no detectable network traffic during our captures. An examination of the domains we detected in network analysis reveals that some of the libraries shown in Table 3.2 were not contacted. This was likely due to our limited ability to conduct dynamic analysis; we could not get past the account login page on these apps and might not have reached the page(s) that made requests to these libraries.

Nearly all of the traffic to and from these servers was encrypted using TLS (aside from one font request made by one app) Additionally, based on an IP address-based geolocation tool [197] we used (which are typically accurate on the country-level [171]), all of the servers contacted were located in the United States.

What data do the apps send, and how frequently are they sent? Using mitmproxy [83] we were able to intercept and decrypt network traffic being sent to some of the apps and observe their contents in clear text. When attempting MITM, two of the monitoring apps (`Telmate Guardian` and `aCheck`) displayed notices indicating they detected the device was rooted and would not permit login. While `Telmate Guardian` contacted the same domains as it did on an unrooted device, `aCheck` did not reach out to its server (gwusacheck.aware.attentigroup.com) when the device was rooted.

Although the volume of network traffic was low, apps sent general device information using both event- and time-driven requests Generally, most apps did not send much data, but those that did sent general device info.

We observed two apps sending time-driven requests. `Telmate Guardian` pinged New Relic once every minute, sending device data and information about domains recently contacted by the app. `Sprokit` contacted Facebook once every five minutes, sending several data. `Telmate Guardian` and `TRACKphone Lite` both sent event-driven requests to third-party libraries (Flurry and Branch, respectively) whenever the apps were moved between the foreground and the background (e.g., if the person using the phone was looking at the app or not). This could be used to calculate the total amount of time someone is using an app or potentially to ensure someone is looking at the app.

A passive observer on the same Wi-Fi network or an entity such as an ISP may be able to know that the person connected to their network is under EM and using one of these apps based on the domains they observe Six of the 16 apps (`aCheck`, `BI SmartLINK`, `Community Supervision`, `IntelliTrack Mobile`, `RePath`, and `Telmate Guardian`) contacted domains that might uniquely identify the apps, meaning that the domains often included the names of the apps or their parent companies; a list of these domains is available in the appendix. For example, `Telmate Guardian` contacted domain `api.telmateguardian.com`. This information could allow passive observers—e.g., coffee shops, airports, schools, employers, Airbnb hosts—to know if someone is under EM.

Concerns in the App Reviews Because the data collection and sharing practices of the apps ultimately impact the people required to use them, it is important to understand what concerns they have about these practices—providing more depth to our understanding of people's actual experiences with these apps in ways that a strictly technical analysis cannot.

To understand the concerns of people using these apps, we qualitatively coded reviews of these apps in the Google Play Store. Twelve of the 16 apps had visible reviews (N = 257) in the Google Play Store; `aCheck`, `Sprokit`, `TRACKphone Lite`, and `Uptrust` were the

exceptions. Below I include the most insightful results from the app review analysis.

Malfunctions discussed by the reviewers of were mostly related to an inability to use the app to successfully perform a check-in—an important requirement of community supervision. This inability to check-in was often attributed to failures in the apps' facial recognition, voice recognition, or location detection systems Some failed check-ins were attributed to general lack of functionality (e.g., R189: "*app [won't] let me check in,it has been having problems all day today*") or not receiving notifications that a check-in was being requested (e.g., R32: "*Does not do notifications. Causes of problem with PO [probation/parole officer]*").

Some reviews (n=33) also mentioned failures that involved smartphone sensors (e.g., camera, microphone, location). Several apps require people to send a picture, send a video, or hold the phone to their face while facial recognition happens. Common problems related to camera usage included inability to take a picture or record a video and not being recognized by facial recognition algorithms. Regarding facial recognition, R37 wrote "*The facial recognition needs to be refined since I didn't have makeup on when I took the first pictures, however when I put on makeup, facial recognition becomes much harder, even in adequate lighting*," and R41 said "*Facial recognition is terrible. I've given up.*" Research by Buolamwini and Gebru [61] showed that facial analysis algorithms have significantly higher error rates on darker-skinned people; this could cause facial recognition problems to disproportionately impact Black (or other darker-skinned) people under EM. A review added by R93 during the international COVID-19 pandemic read "*Can be very inconvenient when I am out in public and have to take my mask off to check in ...*"

R198 expressed frustration with the location sensor: "*Hate it.. it goes off for nothing and it supposed to be gps but can't even detect the right location..... STUPID APP...*" Another reviewer (R150) described their troubles with using the microphone for voice recognition: "*It keeps locking up. I have never gotten past the voice analysis. It truly sucks.*"

Disruptions caused by the apps included 1) loud alerts in inappropriate settings (e.g., work or church) or at inappropriate times (e.g., they were asleep), 2) taking up significant resources on their smartphones, such as space and battery, and 3) causing the entire smartphone to crash or freeze, potentially jeopardizing an EM condition that their phone is always running and available These disruptions could violate the information security principles of availability (if the app causes the phone's OS to crash) and confidentiality (because of privacy leakage when loud notifications happen). The reviews contained descriptions of disruptions they cause in the reviewers' lives, particularly the notifications from the apps and the problems the apps cause on their smartphones. Regarding the volume of the notifications from the app, R129 wrote "*... Raises all media to maximum volume when the notification goes off (even on silent) which is incredibly harmful to your ears with ear buds in.*" Another reviewer (R133) wrote "*... the notification overrides my phone's silent/vibrate function which is a nuisance for certain places (e.g. church, work etc.). When I first started the program I could keep it silent, then it started overriding about a month into it.*" The reviewers also mentioned how the timing of the notifications can be disruptive, sometimes going off while they are asleep—e.g., R184: "*it has costed me a job already because they ping you in the middle of the night while you are asleep. cant wake up on time to get to work*," R180: "*It goes off all night and keeps me awake ...*"

Several reviewers mentioned how the apps they use drain battery from their phones. To remedy this, one reviewer described purchasing an external battery to ensure that their phone always had power. Other complaints were about the amount of memory these apps take up and how using these apps can cause their entire phones to glitch and freeze. Regarding the memory requirements of an app, R228 wrote "*This is b.s. man.. gotta update every few weeks bcuz they keep thinkin of new ways to keep their boots on our necks.. it ends up taking so much space, you gotta buy a phone JUST FOR THIS ONE APP!!!! ZERO STARS.*"

Discussing crashes and battery, R64 wrote "*App constantly crashes in the background, no response from support. Drains battery from constantly reopening and crashing*" and R87 wrote "*Freezing up phone and causes a lot of glitches along with consuming battery life.*"

EM apps were compared to ankle monitors and prisons in a few reviews; EM apps were described as better than prison, and both better and worse than ankle monitors Reviewers of the apps sometimes compared them to other methods of EM, namely ankle monitors. Two reviewers described using an app as better or more tolerable than using an ankle monitor (R91: "*... it's one hundred percent more livable than an ankle bracelet ...*"). Conversely, one reviewer said they preferred using an ankle monitor to using an app, after listing a myriad of problems they encountered with the app—R129: "*... Ridiculous waste of money for something that does nothing but frustrate you. Ankle monitor ¿ [Outreach Smartphone Monitoring] ANY DAY.*" Similarly, another reviewer suggested that people under EM should use a different method for location tracking if they have other options available to them—R179: "*It's a horrible app and if you have a choice of some other gps options take it.*".

Reviewers also compared using the apps to being incarcerated. Unsurprisingly, while they described using the apps as inefficient and dysfunctional, some reviewers still thought using the apps was much better than being in prison or jail. R187 wrote "*Horrible app. Constant network problems, won't let me pay on the app ... extremely inefficient all around. but... it is better [than] prison...*"

Reviewers described a general sense of injustice by being required to use these apps. They also raised privacy concerns, and felt that using this apps would lead to more problems with their EM supervisors and potentially imprisonment Some reviewers (n=9) explicitly mentioned surveillance or privacy concerns that they have with the apps they were using. For example R217 wrote "*It's ok. I don't trust it because it is very intrusive but have no choice in the matter because I am on state probation.*"

Reviewers (n=23) raised the risk of them getting a violation because of the app malfunctioning. R37 wrote "*I've been having trouble with the check-ins not alerting my phone which causes my probation officer to call and threaten to file a warrant for my arrest because I missed the check-ins, which is incredibly frustrating and distressing.*" Similarly, R192 said

"*This app has a very bad default in it ... when trying to report to your parole officer it has a tendency to not allow you to report ... when it says that you need to report it is not logging it in so therefore if you have this app you are going to go to jail because it's going to make you fail ...*"

More generally, 57 reviews mentioned a broad sense of injustice or being wronged. One reviewer (R209), who used an app as part of a drug treatment program, pleaded with the app's developers to fix its problems:

> "l'm a drug court client in phase 5 been in the program over a year done very well[,] worried about this app it doesn't work not very well[,] the developer's should be ashamed of themselves[,] this is my sobriety and freedom that's at stake this app has the ability to destroy all l have work so hard for[,] please fix it or take it down[,] your money is not worth my freedom !!!!"

3.5 Privacy policy analysis

We inspected the privacy policies of these apps 1) to understand their described data collection and sharing practices, 2) to observe what regulatory limits on these practices they raised, and 3) to determine if the behavior we observed during our technical analysis was covered (explicitly or implicitly) by statements in the policies.

Three apps did not have a privacy policy available in the Google Play Store, indicating that they may be in violation of the Google Play Store's user data policies Every app in the Google Play Store "must post a privacy policy in both the designated field in Play Console and within the app itself" [142]. Out of the 16 apps we analyzed, three of them (`Community Supervision`, `Corrisoft AIR Check-In`, `Sprokit`) had links that did not actually point to a privacy policy. `Corrisoft AIR Check-In` even had the words "Privacy Policy" on its website but there was no hyperlink to click on [82]. This means these three apps appear to be in violation of the Google Play Store's user data policies [142].

Only 9 apps had a privacy policy that explicitly addresses the apps' usage, and one of them may have taken down a relevant privacy policy in response to public scrutiny Although three apps did not have a privacy policy linked the Google Play Store, we were able to find one of the policies on the app's website, bringing the number of policies we found to 14. Of the 14 privacy policies that we were able to locate, four of them do not reference the EM apps, with three of them specifically referring only to the "Site" where the privacy policy was hosted. It could be argued that these four policies also violate Google Play's policy because they do not address their respective apps; however, this violation is less straightforward than the aforementioned one. `BI SmartLINK`'s Google Play page had a link that appeared to be to an app-specific privacy policy (`https://bi.com/products-and-services/bi-smartlink-privacy-policy/`), but the URL forwarded to a generic privacy privacy with no mention of the app (`https://bi.com/privacy/`). Interestingly, as recently as May 2021, the app-specific `BI SmartLINK` URL was active and contained relevant information [150]. Later that month, a report critiquing the app and referencing its privacy policy ("*... SmartLINK's privacy policy indicates that the application can share virtually any information collected through the application, even beyond the scope of the monitoring plan, with the supervising officer*") was published [204]; as of October 2021 that privacy policy is no longer reachable. We reached out to the developers of `BI SmartLINK` (as mentioned in Section 3.3) in early March; while they never replied to our email, on March 18, 2022, they updated their link in the Google Play Store to a privacy policy specific to the app (`https://bi.com/bi-smartlink-privacy/`) [1]. This means that only 9/16 apps had privacy policies that appear to be applicable to their respective apps; we describe these nine apps' policies in more detail below.

3.5.1 Data collection & sharing

While the level of details regarding data collection vary, almost all the apps said that they share data with third-parties, sometimes for marketing or advertising purposes While some apps' privacy policies gave very detailed description of what data

they collected—`RePath` even explicitly mentioned the sensitive permissions requested in the app and provided a use case for each one [102]—other apps were quite vague, with text like "we may require you to provide us with certain personally identifiable information" [182].

Eight of the nine policies had language about sharing data with law enforcement, a supervisor, or a court-based entity. Although these policies state that they will comply with warrants, they also outline other reasons they might share data with one of these entities without a warrant, such as to "protect and defend the rights or property of [the company]" [133]. Eight of the nine policies also described their data sharing practices with third-parties. These practices appeared to be similar across the policies; one's personal data is typically shared with affiliates, subsidiaries of the companies, or a service provider. The service providers include companies that do web hosting, marketing, analytics, and advertising.

Regarding selling data, five of the policies said explicitly that they do not sell one's data. Seven of the policies mention that data will be used for marketing, sometimes for marketing the company's own product and advertisements.

3.5.2 Regulations

Apps mentioned regulations but may consider themselves exempt from complying with certain portions of them The privacy policies may be relevant if people under EM bring legal challenges against third-party data disclosures or retention. For example, the California Consumer Privacy Act (CCPA) created certain rights for California residents to request deletion of personal information by private business and permits civil penalties for violations [2]. However, the Act only applies to companies of a certain size or revenue [3], and it is unclear whether the businesses producing EM apps would qualify. Additionally, prior research on prison technology companies indicates CCPA may have little impact, even if it applies, due to broad exceptions within privacy policies [223].

Five policies mentioned CCPA, and four mentioned Children's Online Privacy Protection Act (COPPA). One app (`IntelliTrack Mobile`) described itself as exempt from CCPA's

data deletion clause and that the community supervisor should be contacted, saying it is "generally exempt from the Right to Delete requirements of CCPA. All Right to Know, Right to Receive and Right to Delete requests should be directed to your corresponding Supervising Authority" [133]. Another policy mentioned that "monitored users" may have limited deletion rights [181]. Similarly, while five policies contained text about data deletion and retention, only one (`Shadowtrack`) named a fixed duration after which data would be deleted: "All facial recognition data is stored for a period of **up to seven years** after the Enrollee is removed from the program. This retention time period is dictated by the supervising agency" [268].

Two app policies included the possibility that data might be stored or processed outside of the U.S., bringing into question how the privacy laws abroad may conflict with those of the U.S. and if that affects the monitored individual's data rights.

3.6 Putting our results in legal context

To understand the legal ramifications of these apps and what protections exist for people on EM, we examined the legal context for EM apps. The Constitution and its interpretation by the Supreme Court set the baseline of protection against invasive community supervision practices. Existing protections often arise from *legal challenges* alleging unconstitutional practices; these rulings, if favorable for the people on EM, can set limits on the scope of certain invasive practices.

However, legal challenges to EM of individuals under supervision face three significant hurdles. First, constitutional precedent is unfavorable, particularly when "public safety" is balanced against the privacy rights of a disfavored group like people convicted or accused of a crime. Second, individuals under supervision are already subject to strict conditions infringing the right to privacy, freedom of speech, and religion; arguably, smartphone EM is no different [146]. Third, advocates may be hesitant to challenge EM because they believe its alternative would be greater incarceration, rather than abolition; we discuss this third point in more detail in Section 3.7.3.

Courts disagree on the limits of continuous EM of supervised individuals, and the Supreme Court has yet to decide the issue [300]. The most relevant constitutional protection against government EM is the Fourth Amendment right to be free from unreasonable searches and seizures. Smartphone apps present a search of phone data as well as location data [17]. "The touchstone of the Fourth Amendment is reasonableness" based on the degree of intrusion into an individual's privacy [15]. The Supreme Court, however, has held that probationers and parolees have a diminished expectation of privacy [16], since criminal convictions necessarily "curtail an offender's freedoms" [15]—although, as previously discussed, not everyone under EM has been convicted of a crime (e.g., people on pretrial release or release from immigrant detention). The court balances this diminished expectation of privacy against government interests that include "integrating probationers back into the community, combating recidivism, and protecting potential victims" [16]. Against such vague state interests, "searches are almost always deemed reasonable" [283].

The breadth and continuous nature of smartphone surveillance raises the question: **at some point, surely *some* kind of search must be unreasonable?** However, the government may circumvent the reasonableness requirement altogether by invoking a questionable notion of consent. Some circuit courts have held that because the criminal defendant agreed to warrantless searches in their supervision conditions to avoid incarceration, they have consented to the search and forfeit the right to raise a Fourth Amendment claim [283]—regardless of the obvious issues of coercion (i.e., if you must use the app or go to jail) [146]. Notably, the relevant consent in such a case is not to the privacy policy of the smartphone application, but to the conditions of supervision imposed by a court or administrative body [18]. As a result of both the Fourth Amendment "reasonableness" analysis and consent arguments, groundbreaking Supreme Court cases such as *Riley v. California* and *United States v. Carpenter*, which imposed greater protections on smartphone searches or location data [17, 19], have generally not been applied to individuals on probation or parole [283].

Private companies may face legal challenges as well. Although the Fourth Amendment generally only applies to government actors [14], it may also apply to private actors who are

sufficiently involved in a public action such as administering criminal punishment [110, 241]. Constitutional law aside, private contractors are still subject to statutory, contractual, and regulatory requirements as well as tort law [202], all of which can be used to challenge faulty monitors [107, 121].

Advocates face an uphill battle in distinguishing smartphone EM from other conditions of supervision that have been deemed legally permissible. For example, one district court, in responding to an ICE detainee's claim that 24/7 smartphone monitoring encroached on his individual liberty, noted "far more onerous" conditions such as mandatory lifetime sex offender registry or prohibiting a parolee from leaving the state for four years are legally valid; EM seems tame in comparison [23].

Ultimately, it is difficult for anyone to bring legal challenges if they are faced with a false choice of opting in to EM when the alternative is incarceration. It is necessary to move beyond "alternatives," as discussed in Section 3.7.3.

3.7 Discussion and conclusion

3.7.1 EM apps introduce new harms & risks

Due to their multipurpose use, EM smartphone apps introduce new risks to people, relative to both typical community supervision and ankle monitors. Because of the range of mobile OS permissions and sensors on smartphones, apps can collect and share significantly more data than ankle monitors (even ones that may have microphones). These data can be shared digitally with third-parties used within the app but also can be shared by EM supervisors with police or other prosecutorial entities. This increase in surveillance capability further exacerbates the power imbalance between people under EM and their supervisors.

Using EM apps also means that entities that might not otherwise be aware that an individual is under EM now know this. For example, network service providers (e.g., ISPs) observing network traffic may be able to telling that someone is using an EM app based on a domain. Mobile operating systems (e.g., iOS, Android) log whenever someone downloads

these apps. Third-party libraries (e.g., Facebook) can learn that an individual is using an app, and they may have additional personal data about this individual.

As discussed in Section 3.4, the apps are often unreliable and dysfunctional. Many reviews discussed a variety of malfunctions within the app. Other reviews described how the apps affected the performance of their entire device, causing it to crash. These issues could cause people using EM apps to be more likely to fail a check-in; some reviews mentioned how people felt that these apps were setting them up to fail. Given that people might not be able to successfully check-in, and they need to have their devices on and charged as a condition of EM, it can be argued that the use of the apps is likely to increase interactions between people under EM and their supervisor and increase the likelihood that they might be incarcerated.

Our results indicate that some apps request permissions that let them access audio and video data, and may periodically send data to servers (e.g., `Sprokit` contacting Facebook every five minutes). The amount of data sent by these apps may create financial burdens for people using them. Given that the poorest people are overrepresented in community supervision [157], and poorer people are more likely to use prepaid phone plans [226], apps that send significant video or audio data for biometric verification could be costly. The cost of mobile data plans necessary to use the apps can be burdensome, especially in addition to the fees that may be required as an EM condition [113, 284]. Although the requirement to pay supervision fees was already part of some individuals' community supervision and reflects a broader power imbalance in which the government compels people to be surveilled and then makes them pay for their own surveillance, EM apps can exacerbate the financial burden on people being monitored additional fees.

It is worth noting that while we believe EM smartphone apps introduce new risks and harms to people on community supervision, we do not seek to imply that these new risks and harms are more significant or important than those already imposed by the conditions of community supervision (the existing power imbalance). The restrictive conditions of community supervision (and the predictable failures to comply with these stringent conditions [93]) "can

lead to job loss, housing instability, difficulty caring for children, interruptions in healthcare, and a host of other collateral consequences" [158, 228, 285]. Using EM apps adds to an already onerous list of things people under community supervision must manage, showing how technology exacerbates the power imbalance.

3.7.2 Examining the "least privileged" app

We observed a potentially wide discrepancy along multiple dimensions between the least privileged (`Uptrust`) and the most privileged apps (`Sprokit`). This discrepancy raises the question: what permissions are necessary for these apps to accomplish their goals? In computer security & privacy, the principle of least privilege states that a system "should operate using the least set of privileges necessary to complete the job" [246]; this principle has relevance in this ecosystem. If an app's goal is rehabilitation, it may focus more on features like court date reminders and request the minimum permissions necessary to support those features. If its goal is surveillance, it will likely request as many permissions as possible to collect the widest range of data, and may share that data widely.

That said, we acknowledge that although the apps can be used for similar purposes, they may have different goals (e.g., enforcing a home curfew versus daily breathalyzer readings) and consequently need different permissions. There is no existing standard for what functionality EM apps should or should not include nor what permissions EM apps should request, and from a legal perspective there is no burden (that we are aware of) on the government to choose a less privileged app over one that might be more privileged. Nonetheless, it is still helpful for us to identify the least privileged app to inform policymakers who can develop regulatory limits for EM apps and may use the least privileged permissions access as a model. Our work provides empirical data for a multi-stakeholder conversation to potentially develop a model to determine what permissions are necessary and how much data collection is "too much" (if these apps are to continue to be used).

3.7.3 Moving beyond "alternatives"

As we note in Section 3.6, legal precedent is not favorable to parolees and probationers, or to challenges regarding consent, since people under EM may "agree" to the conditions to avoid incarceration. Since they consent to the general conditions of EM rather than the terms and conditions of an app, people under EM may be subject to whatever data practices the app vendor itself desires (as long as these practices are not disagreeable to the EM supervisor and do not violate contract terms). EM app vendors can force updates and change their privacy policies (if they have one) at any time, and the people coerced to use these apps may not have a successful pathway to legally challenging any of its practices because they "consented."

Regardless of what legal arguments may be raised to challenge EM, it is important to know that the choice is not just "EM or incarceration," as judges and prosecutors may present it; there is also effective community-based rehabilitation. As Chaz Arnett has noted, "the narrow comparison between jail and electronic monitoring" provides an incomplete choice when a variety of abolitionist alternatives may be explored [42]. Most importantly, as Kate Weisburd wrote, "[t]here is no empirical evidence ... that monitoring is used as an alternative; and that in a world without monitors the same people would (or should) remain incarcerated" [42]; a recent report examining pretrial electronic monitoring in Los Angeles County supports this [278].

3.7.4 Recommendations

In light of our findings, we direct our recommendations to technology companies and the computer security and privacy research community.

Mobile app marketplaces Mobile app marketplaces (e.g., the Google Play Store) should realize that they are not neutral actors and that they have a place in the future of EM apps. They can enforce their terms of use and require apps that collect sensitive data to have a privacy policy that describes how the app functions or be taken down from the

marketplace; this could cause up to several apps in this study to be removed. However if they are immediately removed, people who are required to use them or to newly enroll may be unable to do so and may face immediate harm as a result. A similar risk exists if a company removes their app from the Google Play Store (like `Sprokit`) while it continues to be used. EM apps being removed from app marketplaces could lead to supervisors sideloading [265] these apps onto people's devices (instead of downloading them from app marketplaces), and this ecosystem would become more opaque. However, another possibility is that the usage of EM apps would become untenable; the labor required from EM supervisors (managing app updates and complaints) might lead to a decline in their usage.

Relatedly, app marketplaces could also modify their terms of use to limit the use of apps in their marketplace in carceral contexts. App marketplaces could have special rules for EM apps. Just as incarcerated people and people on probation/parole may be considered a "special population" by an IRB, one could imagine a flag that app developers are required to set if their app is used for electronic monitoring. This flag could trigger additional rules, including increased transparency requirements. The Google Play Store already prohibits apps that block ads and apps that allow people to cheat at games [125]. They could similarly prohibit EM apps. In allowing EM apps and banning others, they are making a set of value judgments; our work calls on them to consider whether these value judgments are appropriate.

Computer security & privacy researchers EM apps exist within a broader ecosystem of carceral technologies. This ecosystem includes technologies like ankle monitors, recidivism risk-assessment tools, and mental health prediction targeted at incarcerated people. These same technologies are often administered by private companies on behalf of public sector entities, meaning that they may not be subject to the same public records requirements as governments. Despite the severe impact that these technologies may have on people affected by them, many of their internal parameters and controls are unknown. While understanding the inner workings of these technologies is not necessary to understand the harm they may

cause, it may benefit the public by exposing faulty or discriminatory inputs and the harms that they *do* cause. Given our skills for understanding complex systems and frequently interdisciplinary methods, the computer security & privacy research community is particularly well-positioned to make a positive impact in this space by increasing transparency and, consequently, accountability.

Future work in this space could determine how to more thoroughly study EM apps and overcome the significant ethical concerns and structural challenges. To actually understand usage one needs to have an app that is paired with an account run by an EM supervisor; we do not currently have the structures in place to conduct these experiments.

Other recommendations Regarding recommendations for legislators, judges, prosecutors, state and county community corrections organizations, or activists, we will defer to the recommendations of those organizations and people actively working in these sectors. We refer the readers to the work of Kate Weisburd [283, 284], Just Futures Law & Mijente [204], and James Kilgore [166, 167] for detailed recommendations for each of these actors.

3.7.5 Conclusion

This chapter presented our analysis of 16 Android apps used for electronic monitoring. We found that these apps include numerous trackers, the permissions requested by them vary widely (with the most common one being location), and reviews indicate that their users find them invasive and frequently dysfunctional. This is the first work to systematically analyze apps in this ecosystem that desperately needs transparency and accountability. Our results call for all stakeholders (including app stores, security researchers, and legal stakeholders) to rethink what, if anything, these apps should look like.

3.8 Technology as an amplifier

The sensitive data that smartphones have access to and the blanket consent that users are coerced to give means that people using electronic monitoring apps are subjected to even

more surveillance that they otherwise might be under community supervision. Moreover, the additional conditions that EM apps impose on their release make staying outside of detention and avoiding technical violations even more challenging. It can be argued that EM apps are serving to make a punishing post-release reality for people even more punitive [284].

The reviews we analyzed indicated that some of these apps may cause people's devices to malfunction and freeze or crash its operating system. This violates the requirement that the user be constantly available for a check-in via their phone. In this light, these apps may be increasing the likelihood that people violate the conditions of their release (relative to community supervision).

A parallel observation from our work is that the presence of tracking/advertising domains may mean that EM app developers may profit financially from people coerced to use their apps—on top of the money they make from their vendors (i.e., a county). This mirrors a broader power imbalance in which the U.S. carceral system is extracting wealth from the people ensnared in it and their families [248].

Lastly, we found that entities that can passively observe network traffic (e.g., coffee shops, airports, schools, employers, Airbnb hosts) may be able to identify that someone is using an EM app based on the unique domains called by these apps. Prior work has discussed the stigma (and physical pain) associated with ankle monitors [168], and EM apps have been pitched as less cumbersome and embarrassing and a way avoid this stigma [269]. However, this stigma is typically limited to judgment from others *who can physically view* the person wearing the ankle monitor (either in-person or via a photo). The usage of an EM app facilitates surveillance on a larger and broader scale than previously possible with ankle monitors. People coerced to use an EM app maybe not face stigma from in-person interactions (unless they are actively doing a check-in), but they can be stigmatized, labeled, and targeted by data-brokers, advertisers, or other parties in ways that can harm them.

In Chapter 4 we build on this work and focus on the largest usage of electronic monitoring apps (i.e., immigration surveillance) and its impact on migrants. Migrants experience a different power imbalance than EM app users and consequently have a different threat model.

Chapter 4

UNDERSTANDING EXPERIENCES WITH IMMIGRATION SURVEILLANCE

As we learned in Chapter 3, government entities hold enormous power over laypeople, and electronic monitoring smartphone apps exacerbate that power imbalance. In this chapter, I build on this work to focus on another power imbalance between users and government entities, in the context of immigration surveillance. People attempting to migrate to the U.S. are increasingly subjected to technology-based surveillance after coming into contact with Immigration and Customs Enforcement (ICE). Although other technologies are used to surveil migrants (e.g., ankle monitors), the most common one used is a specific electronic monitoring smartphone app: `BI SmartLINK`. In the following pages, I explain the broader context of ICE's surveillance program, what `BI SmartLINK` is, the risks it poses to migrants (and how these risks differ from those outlined in Chapter 3), and what technologists can do about it. This chapter is based on a paper coauthored with Yael Eiger, Basia Radka, Tadayoshi Kohno, and Franziska Roesner that was accepted for publication at the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT) [224].

4.1 Introduction

Many asylum-seekers (and migrants broadly) come to U.S. borders fleeing various forms of violence or catastrophe and have overcome enormous challenges [84]. Those who are able to gain entry into the U.S. may face additional challenges after they enter the country, including temporary detainment by Customs and Border Protection (CBP), language barriers, the inability to work legally,[1] and finding affordable housing [75]. Moreover, before being

[1]People who have a pending asylum case with the United States Citizenship and Immigration Services (USCIS) may apply for work authorization but only after 180 days [72].

admitted to the country, they must surrender troves of personal data (e.g., through device searches and searches of social media accounts [252]), and, if they are permitted to enter, must accept several conditions—often including being surveilled by a smartphone app: `BI SmartLINK`.

`BI SmartLINK` was first launched in 2018 under Immigration and Customs Enforcement's (ICE) Intensive Supervision Appearance Program (ISAP), which is the primary component of ICE's Alternatives to Detention (ATD) program [138]. According to ICE, the goal of ATD is "to ensure compliance with release conditions and provide important case management services for non-detained noncitizens" [276]. Instead of being detained while their immigration case is pending or released with routine in-person check-ins, migrants are assigned some form of electronic monitoring. These options have expanded over the years to include ankle monitors, home phone voice verification, smart wristbands, and `BI SmartLINK` [207]. According to ICE data from March 2025, of the 183,884 people enrolled in ATD (with an average enrollment of 651 days), at least 159,959 people were being monitored by `BI SmartLINK`, with the remaining people being monitored by an ankle monitor (n=17,689), smart wristband (n=4,634), VoiceID (i.e., phone-based voice verification; n=1,554), or no technology (n=6) [273].

Migrants required to use `BI SmartLINK` must conduct remote check-ins using facial verification and have their location tracked. When this app is installed on a personal phone, it has access to personal data (e.g., images, videos, phone number(s)), in addition to the already sensitive location and biometric data collected. There is a power imbalance between migrants and the entities that monitor them, particularly regarding access to information about `BI SmartLINK`'s behavior. If a migrant cannot successfully complete a check-in, or if they violate the terms of their supervision because of a misunderstanding about how the app functions, it could lead to them being subjected to more stringent surveillance or being detained. As computer science researchers, we seek to understand migrants' experiences with this compulsory surveillance technology, what knowledge gaps or uncertainties they may have about how these technologies function, and what role, if any, we can play in this

ecosystem to make immigration more just and equitable for them.

To this end, we ask the following research questions:

RQ1: *What are people's experiences under ICE's ATD program, particularly with using* `BI SmartLINK`*?*

RQ2: *What are people's questions about the functionality of* `BI SmartLINK`*?*

RQ3: *What are people's recommendations regarding* `BI SmartLINK` *for developers, researchers, and policymakers?*

To answer these questions, we conducted a semi-structured, qualitative interview study with immigrant rights advocates (n=9; see Section 4.3.6 for a discussion of sample size and recruiting) that currently support or have supported people monitored by `BI SmartLINK`. These advocates—some of whom have personal experience with the U.S. immigration system—have collectively supported thousands of migrants under ATD over their careers and can offer unique and valuable insight into migrants' experiences with the surveillance technologies they've been required to use. We asked them about their perceptions of the app, how people have described their experiences with the app, and the questions that they have heard from people that they have supported.

Our results highlight the power imbalance between the migrants and the people surveilling them, the negative impacts of the tech on migrants, and participants' questions about app functionality (arising from a lack of transparency regarding its behavior). Moreover, participants shared higher-level reflections about the pitfalls of viewing increased surveillance as a reform and provided recommendations across the ecosystem, from app developers to policymakers. Our findings regarding the harms experienced by migrants lead us to believe that `BI SmartLINK` should not be used, and these harms cannot be stopped by attempting to improve the app or its design. However, our findings point to potential intervention opportunities for technologists to address transparency around the functionality of `BI SmartLINK` to support migrants, mitigate the power imbalance they face, and inform future policy decisions regarding its deployment.

4.2 Background on surveillance of migrants

4.2.1 ATD, ISAP, `BI SmartLINK`, and critiques

ICE's Alternatives to Detention (ATD) program includes the Intensive Supervision Appearance Program (ISAP). ICE promotes ISAP as a method for reducing absconsion (i.e., fleeing or not attending) from immigration court hearings [276]. According to ICE data, 95% of people under electronic monitoring attended their final immigration hearings, compared to 83% of migrants generally (based on a research report on immigration cases between 2008 and 2018) [99, 274]. This data might suggest that electronic monitoring is an effective tool for preventing absconsion. However, the same report found that 96% of migrants who were represented by lawyers attended all immigration hearings, suggesting that a variety of factors are at play. Moreover, a 2022 report from the U.S. Government Accountability Office (GAO) indicated that ICE presents inaccurate, misleadingly-positive numbers regarding compliance rates and recommends that ICE develops better ways of assessing program performance and provide more oversight of its ATD contractor (BI, Inc.) [275].

Several immigrant rights organizations have previously investigated ICE's surveillance practices [38, 39, 160]. In April 2022 three organizations (Just Futures Law, Mijente Support Committee, and Community Justice Exchange) sued ICE for failing to comply with a September 2021 public records request regarding documents on ICE's use of `BI SmartLINK` [4, 5, 32]. After ICE was compelled to comply with their request, these organizations jointly published a report [160] disclosing their findings. Regarding accuracy of public information, they found that the documents contained information that contradicted claims made by ICE and its parent agency, the Department of Homeland Security (DHS). For example, DHS' Privacy Impact Assessment (PIA) [11], ICE's FAQ [276], and the "ISAP BI SmartLINK Agreement" obtained in the public records request all differ regarding when location data is collected from mobile devices. The report also highlights the role of BI (a government contractor) in making decisions about how migrants are surveilled and how BI is also contractually responsible for helping ICE manage negative publicity regarding ISAP and its usage of `BI SmartLINK`.

Strikingly, the report exposes how ICE conducted a pilot study with `BI SmartLINK` in 2016, found that 56% of facial recognition check-ins failed, yet decided to continue promoting its usage. It should be noted that ICE currently claims that its facial recognition system has an accuracy of 98.5% and has undergone an independent evaluation [276]. Lastly, according the report "in 2017, BI reported that the 'pass rate' for SmartLINK voice biometrics was 75% and that the factors that contribute to the low pass rate 'are not subject to improvement.' " In Section 4.5, we reflect on how the lack of public information and the contradictory nature of this information may increase confusion among migrants and their advocates.

4.2.2 The extended border

Mahmoudi's book "Migrants in the Digital Periphery" highlights the blurring line between borders and border subjects, as biometrics, surveillance, and datafication place the border upon the bodies of the subjects [189]. This foreshadows our findings in Section 4.4.1 that migrants who are not in ATD may not want to live or socialize with those under ATD. In a way, those under ATD carry the border (and its corresponding harms) with them.

4.3 Methods

4.3.1 Author positionality

We are five computer security & privacy researchers trained in qualitative methods, and we have all done research on the security & privacy needs of marginalized populations. Some of us have studied carceral surveillance inside prisons and after people are released from prisons, and its impact on those monitored and their families. Consequently, when we observe power imbalances that are exacerbated by technology, we tend to focus our efforts on understanding and mitigating harm or risks to those most vulnerable.

While one of us immigrated to the U.S. as a child and is now a naturalized citizen, none of us have ever been undocumented in the U.S. or had any interaction with Immigration and Customs Enforcement. Due to our lack of personal connection or experience with migrants

the U.S. immigration system, there may be questions or context that we did not consider when designing our interview protocol. We also acknowledge that computer scientists often attempt to do work that they deem "social good" without adequately engaging with the social context in which relevant social issues arise or considering the negative impacts of their work. While we attempted to be thoughtful and thorough in this work's framing, research questions, and methods, we acknowledge that this may still be insufficient.

4.3.2 Recruitment

To recruit participants we used a variety of methods. We directly contacted local and national immigrant rights organizations to schedule interviews with their staff; some organizations shared our study on national email lists for immigrant rights advocates. We reached out to universities' immigration law clinics. We leveraged snowball sampling [244], allowing people we recruited to connect us with other potential participants. We also made study recruitment fliers and shared digital copies on social media and physical copies in-person at local non-profits.

All participants took a screening survey that asked demographic and contact information and verified that they met our inclusion criteria: that they were immigrant rights advocates who supported people monitored by `BI SmartLINK`. Four participants in our study were attorneys, and five were community organizers. Three participants were formerly undocumented, and another participant was formerly enrolled in ATD. A majority of participants (6/9) were women, and participants fell into three age ranges: 25-34 (1), 35-44 (7), and 55-64 (1). Several were bilingual in English and Spanish. Collectively they supported thousands of migrants over their careers.

4.3.3 Ethical considerations

We considered interviewing migrants currently or recently monitored by `BI SmartLINK` directly and decided that the potential risks to participants were significant. Multiple news reports indicate ICE has targeted migrants who publicly criticize its policies for expedited

removal [9, 62, 159, 240]. Following guidance from Bellini et al. on conducting research with at-risk populations [49], we instead chose to interview *proxies*, namely, immigrant rights advocates that have supported numerous people monitored by `BI SmartLINK`. We sent a draft of our paper [224] to participants to solicit their feedback.

Our study was deemed exempt by our institution's human subjects research review board. We walked through an informed consent document with participants and answered any questions they had before beginning the interview. As the nature of the study might lead participants to mention identifying information about other people, we asked them to use pseudonyms for other people and removed any mentions of people from the interview transcripts.

4.3.4 Interview protocol

All interviews were conducted remotely (between July 2024 and January 2025) and audio-recorded (with participants' consent). We transcribed the interviews using an automated tool [220] that kept audio-recordings and their transcripts stored on our local machine. The interviews lasted an average of 48 minutes (not including the time spent reviewing the consent form), and we compensated participants with a $35 USD VISA gift card that was mailed to them.

We began the interviews with rapport-building questions to help participants feel comfortable. We then prompted participants for any questions they have or the people they have supported have had regarding `BI SmartLINK` and informed them that we would prompt them again at the end of the interview. We then asked several specific questions about the people they have supported; for example, we asked participants, to best of their knowledge, how people start and stop using `BI SmartLINK` and how people use `BI SmartLINK`. We asked participants about what they have heard about people's experiences using `BI SmartLINK` and if they have concerns about people using the app. We closed the interviews by asking participants about feedback that they would give to the apps' developers and about their suggestions for what researchers should focus on. The full interview protocol is available in

Appendix B.

4.3.5 Data analysis

Our qualitative analysis involved inductive and deductive coding. One researcher created an initial codebook based on the first three interviews. Two researchers then independently coded the fourth interview. Lastly, they discussed their coding of the fourth interview and the codebook until they reached consensus on themes and codes, creating several addition codes and consolidating others. The researchers separately coded the remaining interviews, iteratively discussing and updating the codebook as necessary. We include the full codebook in Appendix B.2.

4.3.6 Limitations

In Section 4.3.2 we outlined multiple recruiting techniques we employed; nevertheless, we struggled with recruitment. We leveraged personal contacts within immigrant rights organizations for recruitment. After six months of recruiting efforts, we had nine participants in our study. In three instances, participants suggested that we reach out to other people that we had already interviewed. When we asked participants what we could do to recruit more (such as increasing pay, reducing the duration of the interview, changing our messaging in the recruitment blurb), they shared that people might have concerns about the motivation for the work and concerns that the research would harm migrants. One participant mentioned how they signed up for the study initially but almost did not participate because they were worried that our research would be used to further entrench surveillance in the immigration system. Despite our low number of participants, we reached saturation for our interview protocol, as no new themes emerged in the last several interviews. The average sample size at ACM CHI (a top venue for human-computer interaction research) is 12, and 20% of papers had a sample size of 10 or less [64]. While we hoped to recruit more participants, we are cognizant of avoiding the "threshold myth" [44]: that there is some fixed threshold for participants after which a study becomes valid.

4.4 Results

*4.4.1 Experiences with ATD (**RQ4**)*

While we began our study with intention of focusing on migrants' experiences with `BI SmartLINK`, we found that several of them used the app simultaneously or immediately after being required to use an ankle monitor. Consequently, in this section we describe migrants' experiences with surveillance technology broadly under ATD rather than only with `BI SmartLINK`. Where appropriate and relevant, we highlight the differences between people's experiences with the app versus ankle monitors.

(Abuses of) power imbalances

ICE officials and BI staff (referred to as "case specialists") hold significant power over migrants, including controlling where they can travel, when they have to be at home, which technology is used to surveil them, and whether they are detained in a facility. Conversely, migrants, who are fighting removal proceedings from the U.S., often cannot legally work and experience multiple forms of precarity in their daily lives. This deep power imbalance was highlighted by participants in our study, particularly regarding abuse by officials, a lack of accountabilty for improper behavior, and the role of case specialists.

Our participants describe how migrants they supported experienced power abuses when interacting with their BI case specialists. These abuses of power included lying to migrants about how long they would be in ATD and the steps they would have to take to be removed from ATD (internally referred to by BI as "graduating" from ATD). P7 describes how case specialists have lied to migrants in the past:

> I know they've lied. I know they've lied quite a bit. So at first, like with the ankle monitors, [migrants I've supported] were told, you know, you have to be on it for like, six months or a certain amount of period. And then . . . they return at that . . . six month mark. And then it's just like, an excuse, 'because of this or

that we're not going to remove it.'

P3 similarly notes that "*your case manager might . . . intimate like, you know, [if] you follow the rules, [if] you show up for your hearing, maybe like in six months we'll get you off. But like that's just their word. It's not like you have a piece of paper that you can hold you can hold on to.*" Additionally, migrants described how they were threatened by their case specialists with more technology-based surveillance (e.g., getting an ankle monitor put back on after getting it removed) or detention. As P8 recalled: "*I've heard reports of just like . . . really intense threats of, you know, 'If you don't do this' . . . if someone's complaining about whatever technology they're on, it's like 'well, if you don't do this, we'll take you back to jail, to ICE detention.'*"

Additionally, participants described the apparent lack of accountability of ICE and BI officials for bad or negligent behavior. ICE's website states that its best practice is "to perform compliance reviews every 30 days a participant is enrolled in the ATD ISAP program" to ensure that migrants have "the most appropriate form of case management and supervision" [276]; according to participants in our study, this is not reflected in practice. As P5 states: "*It says that their recommended best practice is to review the case of each person on ISAP every 30 days to make a determination case by case using all the factors about whether they merit being graduated out of the program, terminated from the program. But that doesn't shed much light into what actually happens.*" Migrants may wait for months or years before being unenrolled from ATD either through advocacy by community organization or self-advocacy: "*there's also like clearly a lack of administrative . . . review processes to really evaluate what level of surveillance an individual should experience. Like it really seems like it's just template all of the [same] things for everybody . . . until you complain enough and you come off of certain parts of that surveillance cocktail*" (P7). P8 shared a theory regarding why these reviews and de-escalation of surveillance do not happen as frequently as they should: "*For the people that are compliant, it's pretty rare to see ICE affirmatively following its own policies and affirmatively taking people off of this tech. And I think that*

there's probably some incentives for ICE to keep people on this technology."

In a structural sense, ICE and DHS displayed a lack of accountabilty by failing to release a Privacy Impact Assessment (PIA) for ATD (as required by Section 208 of the E-Government Act [10]) until April 2023—almost two decades after the program started in 2004. This is something that advocates have requested for years, but only got recently, and it is unclear if there were any consequences for such a delay. As P3 points out "*you know that the privacy impact assessment for this program; they operated it for like close to 20 years without one.*"

ICE claims that it provides "case management" for migrants, but participants pointed out how different what BI case specialists actually do is from case management. According to ICE, a goal of ATD is to "provide important case management services for non-detained noncitizens" [276], which implies that case specialists (sometimes referred to as case managers) administer these services. Participants described how case specialists are more akin to a parole or probation officer than a case manager; for example, P8 stated:

> I think that the case manager term is something that's been brought over from social work and is seems to me deliberately confusing or ambiguous ... These are not social service agencies. This is a for-profit, carceral corporation and its primary function is to surveil and coerce the people that are under its control. I think that most of the interactions with these case managers is more like something that I imagine is akin to interfacing with a probation officer. They're not there to help you, they're there to make sure that you're doing what you're told to do.

ICE says that case specialists offer community referrals for different resources (e.g., food, housing, health) to migrants [276], but P3 says *"The reality is they give people like a piece of paper that says like 'call this number for a food pantry.' That's not case management.*"

Migrants' relationship with their case specialist can be complicated, as it can feel like (perhaps accurately) their freedom from detention or unenrollment from ATD may depend

more on their relationship with their case specialist rather than their own actions. Their case specialist can recommend to ICE that they be unenrolled from ATD, but they can also recommend detention. As P7 noted, it can be challenging for migrants to manage this type of relationship.

> But a lot of the time it's a really difficult relationship to navigate in which there are all these really weird power dynamics. And there's a lot of like having to appease your case manager and like keep your [case specialists] happy. And there are a lot of, there's a lot of verbal abuse and a lot of threats that happen between those interactions.

(Negative) impacts on people

Participants described how both technical (e.g., `BI SmartLINK` or an ankle monitor) and non-technical (e.g., home visits) aspects of being enrolled in ATD impacted migrants' lives.

The requirements of an ATD program can cost migrants their jobs. Participants described how the demands of remote `BI SmartLINK` check-ins have led to migrants losing employment. P7 explained that:

> There's also been a lot of loss of employment as a result of SmartLINK because if they ask you to take a picture on a Tuesday at two o'clock and you're in an assembly line at a factory, you can't just leave. So a lot of [my clients] have lost jobs because they don't have the kind of employment where they can excuse themselves to go take a picture in the bathroom for 10 minutes. One [person] lost three jobs as a result of it.

P7 described the impacts of having mandatory home visits that are not at a specific time on a given day, requiring migrants to stay home waiting for the case specialist to visit:

> So a lot of people are losing, that's five days a month of employment for these surveillance requirements when they also have SmartLINK or when they also

> have an ankle monitor. So they have the GPS capability, but they're also being forced to stay home and lose an entire day's worth of work.

Being monitored puts others in harms way, leading to housing difficulties and alienation. Participants noted that their clients often struggled to find housing, as surveillance of themselves would mean the surveillance of those with whom they lived. P1 described how one of their clients lost housing because the other (undocumented) housemates were *"very worried that ICE would come to the door."* P7 also described this, by explaining:

> I know that with the ankle monitor, it has been hard for some people to have a place to stay, especially if they're staying with non-asylum seeking undocumented families. The idea of bringing something that's connected to immigration that has GPS feels scary and harmful. So I know that when it's something as visible as an ankle monitor, it's been a problem with housing.

Beyond housing, P5 described the general alienation clients experienced because they were *"avoiding gathering with any friends or family because they didn't want to endanger their loved ones by putting a target on their back so to speak."* These housing challenges reflect Mahmoudi's observation that for migrants, surveillance "practices of digital bordering go beyond material borders and seep into the realm of the everyday" [189]. Because `BI SmartLINK`'s surveillance practices is not dissimilar from those that happen at borders, in a sense, by using the app migrants *carry the border with them* and separate themselves from loved ones to protect them.

Being monitored impacts migrants' mental and physical health. Participants described how the instability, fear, pressure, and stress while being monitored harmed their clients' mental health. P1 said that their clients live in a *"great deal of just kind of constant anxiety that they're gonna get redetained or that they're gonna get deported."*

P7 said that *"I think there's always the looming threat of if you don't comply then we will physically detain you . . . I've heard of someone having a panic attack and being forced into an ankle monitor. That doesn't sound like consent to me."*

P6 connected this fear and stress to the lack of transparency about what data is actually being collected by the technology. They describe how their clients *"feel monitored all the time. The fact that they don't know exactly what the data is used for is very scary."* But the pain is not just psychological, it can also be physical. P7 explained how their client "*developed really horrific medical symptoms, like both psychological and physical as a result of the ankle monitor."*

Migrants might also experience stress as a result of officials' mistakes and misleading advice. On an individual level, ICE officers and case specialists sometimes are required to manually input information about migrants, such as their address. When this information is entered incorrectly and an alert is triggered (e.g., from being in a location that is not known to be one's home after a curfew), migrants are the ones who are under stress and penalized, not the government or its contractors. Regarding location P1 said: "*I've heard of instances where like the data gets entered kind of wrong somewhere. So like, . . . they're being told that they're out of their region and they're not.*"

Given that ICE and BI administer ATD, migrants might expect that they have fairly accurate information regarding the immigration process and legal proceedings. However, some participants described how migrant sometimes get bad or false legal advice. P2 said, "*ICE will tell them a lot of things that are not true . . . they don't know what they're talking about. So they'll give them legal advice . . . They are, yeah, not to be trusted.*"

Being monitored may impact access to healthcare. Not only does being monitored negatively impact migrants' mental and physical health, but it may also impact their ability to get healthcare. P1 explained that *"you know I also really worry about [them] getting health care like if you have a sick kid and you need to get them to the hospital and it's not in your 100 mile radius or it's after your curfew. I really worry that people might not be inclined to seek out emergency care."*

Passports are confiscated. Participants described how migrants' passports are used as a bargaining chip. Supposedly, if a migrant turns in their passport, they are unenrolled from ATD or have an ankle monitor removed (while still being required to use `BI SmartLINK`).

However, as P7 articulated, *"They'll confiscate them forever. And sometimes they'll take the ankle monitor off. Sometimes they don't. And then they also leverage it as a threat. 'If you don't bring me your passport from your home country, I will then have to put an ankle monitor on you."'* This threat is arguably as harmful as the technology itself, as P7 concludes: *How harmful is this technology, but also how harmful is just the threat of this technology?"* P8 explained how this was likely desirable for immigration officials because it makes for an easier deportation. Unfortunately, passports may also be migrants' sole valid form of ID: *"For ICE to have a valid passport for someone is from their perspective like one step closer to being able to like actually remove someone . . . [and if] they're undocumented, for example, North Carolina passed a law a while ago where you can't use other forms of ID. Your foreign passport is your only legally valid form of ID for school registration, notarizing documents, stuff like that."*

P8 continued:

> [Handing over your passport is] a kind of bargaining chip with a lot of complications for clients who either aren't able to get one in the first place or, if they do hand it over, then they're stuck with another set of complications in their lives.

Given that being placed on monitoring technology is used as a threat to coerce migrants into handing over their passports, P7 connected the dots between technology and the passport confiscation, by describing:

> I don't understand how this is like legal . . . especially recent arrivals, it's their only form of ID. So it's a huge deal that they're being [confiscated], and it's a person, it's private property. It's personal property. So I don't, I don't understand how that's legal for them to take a foreign passport. And I also think that . . . technology is being used as a threat in order to get this private property. So like . . . how harmful is this technology, but also how harmful is just the threat of this technology?

Monitoring opens migrants up to other surveillance. Participants described how `BI SmartLINK` changes migrants' typical privacy behaviors, because they have to allow for the monitoring. P5 explained how their clients were forced to keep location services turned on at all times:

> Well, I've heard that people have not been able to switch off their location settings ... I've heard at least one person say that when they tried to disable the location services on their device ... they were contacted by BI and, you know, told they needed to stay on location services at all times. And I did read the agreement that people are coerced into signing at the time of enrollment into SmartLink and it says that they agree to keep their location services on at all times.

The monitoring technologies have usability problems with significant consequences. Participants described problems with `BI SmartLINK`'s facial recognition software used to confirm the identity of a migrant during a remote check-in. P1 explained that it was common: *"When they try to do the selfies for the facial recognition ... the phone doesn't accept them."* P8 also described how *"the facial recognition technology has been reported to be worse and less accurate in terms of recognizing folks of darker skin tones."*

Participants also mentioned frequent hardware problems with the ankle monitors. P5 described how *"They were running into all kinds of battery failures with their ankle monitors, because they had the interval for location tracking set to continuous, and that would constantly ping the device and drain the battery and their batteries were crappy."* P1 echoed this, explaining that *"On the ankle monitors the batteries go out you know and not recharging. Certainly physical problems with the ankle monitors swelling and that kind of thing."*

Having to use `BI SmartLINK` can create problems for many migrants who are low literacy or speak languages not supported by the app. P8 described how *"It doesn't support more than three or four languages, I think. It doesn't, of course, account for people that have limited literacy or no literacy."*

Another purported function of the app is to remind migrants of upcoming appointments and court dates, but participants described how siloed, and often wrong, this data is. As P2 expressed: *"But like, the [Executive Office for Immigration Review], which is like the immigration court system, doesn't communicate with ICE. And so [my clients] have to both go to their ICE check ins and then their immigration court hearings. And sometimes they can confuse them and miss one or the other and they're screwed."* P1 explained that:

> The app doesn't tell you about your [immigration] hearing dates and so . . . as a migrant you think you're [good] checking in on the app. You're going to your [in-person] reporting requirements . . . But then the actual immigration court hearing dates that are the most important dates of all, they don't tell you. And they change all the time.

Altogether, these problems cause great stress. These usability problems carry tremendous weight because technology failure can lead to the detention of a migrant who is accused of not being available or present for a check-in. P1 summarized this by describing:

> I've worked with clients who are having to . . . leave a meeting, an important meeting with me, because they got to go deal with the check-in and then it doesn't work and then they're freaked out because they think they're about to get arrested. And I mean it's . . . like all the time, you know, you're worried about this stupid app.

This is particularly frustrating because ostensibly the app is about generating and increasing compliance with immigration mandates, but the tech itself leads to glitches that prevent compliance. P2 describes:

> [My clients have] concerns that it was maybe not functioning or like making a weird noise or just concerns about compliance and then kind of same thing with the phone app check in, concerns that maybe their phone broke or they dropped it and worries about actually complying with the check-ins themselves.

ATD's promises fall short in practice

Participants described how their perspectives regarding the potential positive impacts of ATD have changed over time.

Advocates originally supported ATD because they thought it would genuinely be an alternative to detention. P3 explains:

> We all kind of bought into it too right. Like if you look at some of my organizations and other organizations in the early 2000s, we supported the ATD program. We wanted funding for the ATD program. We really thought it would be the way to get rid of detention but now . . . it's become pretty apparent that like we kind of have to start from scratch.

P1 echoes this sentiment:

> I am really not happy with the use of this sort of technology. I think it's really invasive and awful, and I feel like initially in the advocacy community we maybe didn't realize how bad it was and so people were like 'oh don't detain people just put them on an ankle monitor.' and I think we've now mostly realized that that's a bad trade-off because it should be neither.

ATD is not a real alternative. Despite initial optimism, participants now believe that the Alternatives to Detention program is not a true alternative to detention, but rather *"digital detention"* (P6).

Expanding on this, P6 explains how:

> Also it's not a kind of endpoint to anything. You can be put back in detention anytime. You can be, even if you somehow get off the program, you can be put back on ATD anytime. . . the main goal of the ATD program is actually to show up to court, compliance for that. So it's really just surveillance, right? It's very intense surveillance.

In addition to possibly being *"put back in detention anytime"*, P6 also explained how the expectation of ATD replacing detention or being the precursor to freedom is misleading. They describe:

> Again, a lot of people think that it's a trajectory [where] you go from physical detention to ankle bracelets to phone app, and then [no detention]. And that's not the actual trajectory, the actual trajectory is that you go to the court. This is all for making sure that you comply with court orders to appear, and then you're able to be deported. But I think the way that it's talked about, it's called alternatives to detention, they'll often say that it's a way to keep the community with the community. It's just being kind of like, oh, it's a humanitarian solution. But there's I think lots of intentional misinformation about how it ends for people.

P9 echoed the same sentiment: *"I mean one of the things they sell you on . . . it's just like ATD isn't an alternative to detention, that's bullshit . . . it's imprisonment for sure. It's just open open air and a different kind."*

This sentiment echoes work by Sarah Sherman-Stokes: rather than being a real "alternative to detention," surveillance technologies administered by ATD create "[d]igital cages, masquerading as a more palatable version of enforcement and surveillance, [which] create devastating harms that are hidden in plain sight, while duping us into thinking of these measures as more humane" [255].

P9 described how the technology is *"a way to control and contain and in some ways I think SmartLINK . . . I do think it's a way to kill people,"* particularly because the technology can facilitate and increase the efficiency of deportation, which means many migrants will be forced to return to the violence that they fled.

ATD has expanded beyond its stated goals. Participants described how ATD has become the default for all migrants and not only those who meet certain restrictions. P3 describes the expansion of the program: *"The numbers that we see now for the ATD program are huge compared to what it used to be in the early 2000s. The clients that I saw that had*

it, it was not the norm across all of them." They additionally report that, unlike today when being released from physical detention likely results in electronic monitoring, *"when I was practicing 10 years ago, it was not unusual for somebody to be released from government custody without any restrictions."* P6 also describes the expansion of ATD, particularly during COVID. They explain:

> During COVID, ATD went up, especially because we shut down three of the four detention centers in [my state]. And especially during COVID, the ATD numbers went up like 274%. And my sense has been that ATD has brought more women under surveillance, that has been one of the big outcomes. Because there used to be men who were detained or even put on ankle bracelets, but at this point, I think lots of women are on the activity program.

Although ATD's stated purpose is about increasing compliance for court requirements, participants felt that this is no longer—or was never—the case in practice. P1 expressed that:

> It's not actually about making sure that people appear for all their hearings because in the vast majority of cases they have [ankle monitors and other technology] taken off at some point before their final hearing . . . [which is the] period when you'd want to make sure that they're actually going to come to their hearing and that if they get ordered deported that they're locatable and all of that. But that's not the way it works. It's almost always on the front end just kind of arbitrarily slapped on and then at some point arbitrarily taken off. It would actually be much more logical if you're going to use it—and I am not promoting its use at all—but it'd be much more logical to use it like later in the process like in the months leading up to your final hearing because that's the period when you know you'd want to make sure that they're actually going to come to their hearing and that if they get ordered deported that they're locatable and all of

> that. But that's not the way it works. It's almost always on the front end just kind of arbitrarily slapped on and then at some point arbitrarily taken off.

P8 also described how, if compliance was truly the purpose of the program, there are much more effective, less harmful, solutions. They explain:

> There are studies that document that providing legal representation to people is at least if not more effective in making sure that they comply with going to every court day and complying with the law. There are other alternatives looking at community-based case management programs with actual social service agencies that would help stabilize folks who have recently arrived or who are coming out of [detention] providing actual resources. Those are the things that help people do what they have to do under the law.

ATD actually hurts compliance. Despite the stated goals of increasing compliance, participants described how the current ATD conditions actually work against this goal. P3 described how: *"people just get fed up and they cut off the ankle monitor or they get rid of the app because it's so onerous"*. P9 described how following the rules is an undesirable path because it opens you up to further tracking, surveillance, and subjugation. They explained:

> The more you do that stuff, the more they're getting information on you, they know where you are, they're controlling everything, but that's also the route toward securing asylum. So it's this very tricky thing because on the one hand you want to follow the ISAP rules because ... you know, you're doing all your stuff, you're trying to just be perfect.

They further explained how the trauma and violence experienced in detention after a legal border crossing encourages a self-preservation instinct to cross illegally and avoid experiencing that harms of further surveillance. P9 describes:

> You know people tend to think about undocumented people and they tend to think about crossing the border and the dangers of the desert and all that. Going

> through a legal route to request asylum at a port of entry, doing everything right, unfortunately is more life-threatening, I think . . . I've never met anyone actually who's gone through the detention experience of the border without experiencing some form of torture. Either through being put into what they call an 'ice box' which is a very cold room like 50 degrees or . . . having . . . very bright lights on all the time, 24/7. Not being given access to medical care . . . physically assaulted, sexually assaulted . . . Being separated from their children or spouses is common, being humiliated is common by the guards. And then when people are released they're released with a tracking device, well now they know where you are. And if you're undocumented and you cross the border [secretly] . . . they don't know where you are.

While this sentiment from P9 describes how people experience psychological and physical torture [40] while in ICE detention, scholars have outlined how carceral surveillance technologies (including electronic monitoring apps like `BI SmartLINK`) extend psychological torture outside of carceral spaces into other places, such as migrants' homes. [213, 251, 267].

DHS is moving towards a future where no one is unmonitored. Participants predicted that this surveillance will likely grow to surveil *everyone* who comes through government custody. P7 expressed that: *"I think that these types of technologies will probably just increase over time and I'm really concerned about where we're heading."* P3 likewise stated: *"DHS . . . is working towards a place where nobody is released without any sort, like everyone released from government custody is going to be subject to some level of supervision."*

*4.4.2 Questions about `BI SmartLINK` (**RQ1**)*

We explicitly asked participants if they had any questions or areas where they would like clarification regarding `BI SmartLINK` and its behavior or functionality. These questions or knowledge gaps regarding `BI SmartLINK` have an impact on migrants' lived experiences

when interacting with this technology. The below results highlight potential opportunities for researchers to technically investigate the answers to these questions and offer increased transparency for migrants and their advocates.

The most prevalent question, raised by every participant, was about the nature of location tracking: "*I am obsessed with figuring out the extent to which the Smart Link app can continuously track the geolocation of individuals*" (P5). On ICE's FAQ page for ATD [276], it states that "*BI SmartLINK® is **not** capable of persistent tracking when loaded on a participant provided device*," and that while it is possible on BI-provided phones, ICE does not use this capability. Participants were aware that ICE says that it only tracks migrants' location while they are actively using the app. However, based on anecdotal experiences and news exposés, participants questioned if this was true. P3 said:

> ICE says that they only track someone's location when the app is being used but . . . I keep hearing from individuals who . . . they've received phone calls from their case manager from the ICE officer in charge of their case asking them like, why they were at a place, at a certain place. And clearly the only way they could have figured that out is that they were tracking . . . them on their phone.

Participants also asked questions about `BI SmartLINK`'s behavior regarding data collection, sharing, use, retention, and storage. A prevalent question was about the app's ability to collect other, non-location information from a migrant's smartphone such as their contacts, stored photos, or activity on other apps. Participants were unsure about the potential scope of data collection and wanted to understand what was possible. For example, P4 asked "*what is the data being collected from your ability to track me and . . . what are you accessing on my phone that I'm not aware of?*" Participants' questions regarding data sharing & use highlighted concerns about data being shared outside of ICE and that data being used to detain not only people monitored by `BI SmartLINK` but others around them.

> I think that there's reasons to be concerned about what ICE would do with that

> data, not just under [the Biden] administration, but under especially a more overtly hostile anti-immigrant administration. I would not be surprised to see similar kinds of raids to what occurred in 2019 in Mississippi.[2] I mean, Trump is promising mass deportations and this data would help ICE in a very granular way locate not just these people, but of course these folks are embedded in communities with lots of other immigrants, lots of other mixed status families. And so it would bring ICE to their doors pretty quickly. So, yeah, what is ICE doing with the data? (P8)

Some other questions were about where collected data was being stored, how long it was retained after someone is unenrolled from ATD, and whether their data was shared with third-parties and private companies. These questions highlight participants' concern about "function creep" [55]—when data originally collected for one purpose is used for another—and the ways that data collected about them might be used to harm migrants and those around them.

Participants also raised non-technical questions on topics beyond the app's behavior. They asked about the legality of passports being collected in exchange for changing the technology used to monitor migrants (Section 4.4.1) and the legal limits of `BI SmartLINK`'s data collection practices. There were several questions regarding ICE's administrative decisions or policies, or why certain people have multiple surveillance mechanisms (e.g., ankle monitor and `BI SmartLINK`) on them at once. One participant wanted to understand the prevalence and frequency of home visits under ATD. Another participant wanted to understand why some people use BI-provided phones and why others do not.

The answers to these questions have important implications for migrants' safety and human rights. If migrants believe they are not being tracked when they are, they may increase risk to other migrants with whom they interact (e.g., by going to a previously-unknown

[2]ICE conducted the largest workplace immigration raid ever in a single state. Seven food plants were raided, and 680 people were arrested. Unsealed court documents revealed these locations were chosen, in part, based on ankle monitor location data [156, 254].

gathering place). Moreover, depending on which data are collected, `BI SmartLINK` could be violating migrants' privacy rights. Although courts in the U.S. have found that certain classes of undocumented immigrants can be denied Fourth Amendment rights (i.e., against unreasonable search and seizure) [217], courts in the EU have found attempts to exclude undocumented immigrants from the General Data Protection Regulation (GDPR) [37] to be unlawful.

*4.4.3 Recommendations from immigration rights advocates (**RQ2**)*

In our interviews, participants provided recommendations for developers, researchers, and policymakers, regarding how to improve the state of ATD technology use.

Recommendations for developers.

When we asked our participants if they have any feedback for the developers of the app, every participant said a variant of what P6 succinctly recommended: *"destroy it"*. P4 asked, *"can you just get rid of it?"*. P7 expressed that *"these apps are a form of social control and they should not exist"*. P2 noted how *"there's all these super like talented, smart people that again use their skills for evil"*, and P8 said:

> Just don't [make it]. Yeah, I don't want this technology to exist. I don't think that it's necessary. I don't think that it's helpful. I think that it is used to extend ICE's reach into people's lives, into the lives of immigrant communities and to put a gentler face on government surveillance and control of people.

P6 expanded on how the original use case for the technology informs its use today:

> I don't think people need to be tracked. As far as I know, the app, BI actually developed it as to like track the movement of cattle first. [3] And in some ways

[3] According to a 2022 press release, BI's "founders used Radio Frequency (RF) technology to create a feed management system for dairy farmers to increase milk production ... In 1977, Judge Jack Love of

> that says it all . . . you know, [the United States is] a place built on people's enslavement, enslaved labor and theft and genocide. So it's the same thing. It's like people just . . . track cattle. Are you using it to track migrants? What does this say?

Short of destroying the app and the technology, P3 suggested that the app should at the very least include *"more transparency over the geo-monitoring and what they can listen in on"* in addition to adding a support line for technical problems: *"the app needs to have a better technical assistance line because people end up having like nightmares that they're about to be hauled off into immigration detention jail because they can't upload [photos to] their phone and there's no like 800 number you can call to quickly get assistance with technical issues."*

Recommendations for researchers.

Participants recommended that researchers focus attention on projects for social good. P2 encouraged researchers to *"support things like mutual aid, collective action, you know, organizing, in general"*, P1 explained that *"we need some real studies of what the mental health impacts are of this kind of monitoring because I think they're real"*, and P2 & P7 recommended researchers study how carceral technologies function and extend surveillance: *"[Researchers should study] the way these technologies function and . . . the way the government or private companies are using them to increase surveillance"* (P2) and *". . . these technologies are incredibly harmful and it just seems like the advocates who are exposed to it the most also don't have the time or necessarily the expertise to figure out how to support the, like the fight against surveillance. So it's like, it's an issue that just like a handful of people end up really spending time on"* (P7).

Albuquerque, New Mexico, read a comic book about a villain that used an EM device to track Spider-Man . . . in 1982 he worked with National Incarceration Monitor and Control Services (NIMCOS) to develop an EM prototype. BI acquired NIMCOS in 1984, and the EM industry was born."

4.4.4 Recommendations for policymakers.

The recommendations for policymakers echo those for developers. P1 expressed *"I don't think we should be using it at all"*, and P4 explains how *"One, I don't think anybody should be placed on any kind of surveillance or monitoring. Two, I just, there's no reason for it. It's, you know, even the rates of like, what they call absconders who, you know, abandon the app and just, you know, go freely, it's so low".* If these policies do continue, participants mentioned reforms like standard timelines for reviewing cases: *"that's one of the key problems with the program now and that is why I believe strongly that the program should have designated benchmarks in which people's cases . . . they know their case will be reviewed."* and furthermore, to reduce high caseloads by lowering the number of people on ATD as opposed to hiring more ICE officers: *"I think it's like one officer for every like 600 cases.* [4] *They're not doing the regular reviews now, I personally don't think the solution is they should hire more ICE officers. The solution should be be much more limited in who you desire to enroll in this level of supervision instead of just giving it to anybody because they are standing in front of you."* (P3). P1 echos this sentiment, the ATD should only be applied to the most serious of cases, and not be default: *"if we are going to use it only in truly serious cases where there is an actual flight risk where the person will really be otherwise detained, then I'd be willing."*

4.5 Discussion and conclusion

4.5.1 Opacity by design

Migrants enrolled in ATD lack transparency into multiple aspects of their surveillance, including how long they'll be monitored, why they received a specific technology assignment, how they can 'deescalate' their monitoring to a different option, how they can be unenrolled from ATD, what the thresholds are for behavior that could result in detainment, and the technical behavior of the technology that monitors them. This opacity, combined with the

[4] According to ICE's website, the caseload is 1:125 [276].

discretion given to BI employees, facilitates perpetual stress and insecurity regarding migrants' futures and freedom, independent of the already stressful nature of their pending immigration removal proceedings. Unfortunately, this lack of transparency does not seem to be the type that can be resolved by better user education (as academic researchers often call for). As we discuss in Section 4.2.1, ICE's internal documents and public statements regarding their practices or how technology functions are contradictory. It is hard for migrants under ATD to make informed decision regarding their behavior while being surveilled when there is not a stable ground truth to inform these decisions.

4.5.2 Electronic monitoring of migrants

Like others under electronic monitoring, migrants bear the mental toll of surveillance and how it "transforms the most private spheres of life—our bodies, homes and families—into highly regulated carceral spaces" [153]. Migrants exist in a precarious situation in which their ability to live, work, and experience community is restricted before they are even coerced to use technologies like `BI SmartLINK`. For migrants who continue to be forced to use `BI SmartLINK` despite complying with everything they are told they need to do to be removed from ATD, it can feel like the surveillance itself is a form of punishment [284].

It is important to acknowledge that while the risks (or the threat model [13]) for migrants is similar to U.S. citizens under electronic monitoring [42, 222, 284], they are distinct in important ways. For example, if someone under electronic monitoring (e.g., via a smartphone app [222]) as a condition of pre-trial release [216] violates a condition of their release (as determined by the app), they might be put in jail/detention, similar to a migrant. And like a migrant, the person on pre-trial release has a pending legal case (in this example it is a criminal case instead of a civil immigration one), the outcome of which they will still be subject to, *even if they use the surveillance app perfectly.* One difference arises when the legal proceedings are resolved. The citizen on pre-trial release will, at worst, be incarcerated. The migrant is at risk of *deportation.* Moreover, the app that the migrant used ("perfectly" in this example) may be used to facilitate their own detention and deportation. Lastly, `BI`

SmartLINK introduces risks (namely detention and deportation) to other migrants that may be around the person monitored in ways that the pre-trial release electronic monitoring app does not. While efforts to challenge the use of carceral technologies in these different contexts are related (e.g., BI SmartLINK is advertised as a general electronic monitoring tool for a variety of domains [12]), the use of technologies like BI SmartLINK distinctly exacerbate the marginalization of migrants.

4.5.3 The role of technologists

Barbaras [46] outlines common mistakes that technologists make when investigating carceral technologies: "(1) 'proving' harm, (2) adopting deficiency narratives and (3) optimizing harmful systems." In our work, we attempt to avoid these mistakes. We do not try to quantitatively prove harm but rather engage with advocates who understand and can relay the minutiae of harms based on their vast experience supporting migrants. Rather than focusing on the shortcomings of migrants and helping them cope, we focus on the existing power imbalance in the U.S. immigration system and how technology exacerbates it.

Researchers may be tempted to assume the failures of BI SmartLINK and technologies like it are questions of implementation, and that its problems are "resolvable through changes to input data and deployment" [164]. We do not aim to improve the functionality, performance, or even privacy within BI SmartLINK. In line with prior work that considers the ethics of conducting usability testing on tools for oppressing undocumented people [47], rather than seeking to improve the functionality or efficiency of BI SmartLINK, we focused on understanding the questions of people monitored by the app with the goal of eventually increasing transparency around its behavior. We have determined that the app is harmful and seek to promote opportunities for technologists to help migrants and their advocates that are not "reformist reforms" [131] that "limit their objectives to the maintenance and practicality of the current system" [232].

4.5.4 The limits of reports

In its July 2022 report on ATD [275], the Government Accountability Office (GAO) called for more oversight of BI by ICE and for developing mechanisms to ensure that it is meeting the demands laid out in its contract. Other recommendations included things like improving the accuracy of its data, improving metrics for tracking if ATD is successful at achieving its stated goals, and ensuring that reviews of migrants' cases (also known as "supervision reviews") are happening at the appropriate cadence (i.e., 30 days as opposed to six months as some ICE officers they interviewed stated). Almost three years later, only one of its ten recommendations has been addressed, and it was that migrants be given access to "legal orientation presentation [275]." This report reveals that the U.S. government is aware of some problems with ATD highlighted in our study (e.g., the frequency of supervision reviews), has published its own reports on these problems, and seems to be largely inactive in working towards resolving them.

When we researchers think about the impact of this work and what we hope it will do, we are soberly aware that yet another "report" may not move the needle towards a more just, equitable, and less surveillance-driven immigration system. We hope that our work can inform the FAccT community about the experiences of those surveilled under ATD and those who advocate for them. By interviewing advocates and soliciting their questions, we aim to lay a foundation for future critical analysis of compulsory immigration surveillance technologies that attempts to answer their questions.

4.5.5 Conclusion

Although technology is only one component of the larger system of surveillance and control in the U.S. immigration system, our work shows how it can exacerbate already challenging circumstances for migrants. Justice and safety for migrants, particularly asylum-seekers, will become more precarious as immigration policy in the U.S. is expected to become more hostile towards migrants [51]. Participants in our study called for researchers to study

how technologies function and how institutions are wielding them for increased surveillance. We embrace this call and hope that this work serves as a synecdoche [27] for the FAccT community and the computer science research community more broadly.

4.6 Technology as an amplifier

Migrating to the U.S. is a difficult process, and technologies like `BI SmartLINK` make it more difficult. The questions we solicited from advocates lay the foundation for future work attempting to mitigate this power imbalance. Increasing transparency by answering questions from migrants and their advocates empowers them to modify their behavior (or not) in light of the information presented. Said differently, it increases their agency, even in a coercive environment. Both this work and the EM app work discussed in Chapter 3 highlight how technology (electronic monitoring apps in this case) exacerbate pre-existing power imbalances (i.e., stemming from community supervision and being a migrant in the U.S., respectively) between users and government entities. In the next chapter, we focus on a different type of power imbalance—namely, between users and corporations.

Chapter 5
DECEPTIVE DESIGN PATTERNS IN VOICE INTERFACES

There has been a long-standing power imbalance between users (aka consumers) and corporations. In fact, consumer protection agencies around the world—e.g., the Federal Trade Commission in the United States, the Federal Competition and Consumer Protection Commission in Nigeria, various national agencies within the European Union—exist to protect consumers from potential negative impacts of this power imbalance, such as "unfair or deceptive acts or practices" [119]. Corporations control the technology that users use, how they can use the technology, and how much control users have over the data that platforms compile about them. They can leverage this control to deceive or manipulate users to spend more money than they intend or choose settings that negatively impact their privacy. In this chapter, I explore the implications of voice interfaces in this power imbalance and how corporations can leverage them (intentionally or not) to harm users. This chapter is based on a paper that I coauthored with Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner, that was presented at the European Symposium on Usable Security in 2022 [225].

5.1 Introduction

User: *Voice Assistant, cancel my subscription.*

Voice Assistant: *To manage your subscription, please visit our website.*

Deceptive and manipulative design patterns (sometimes called "dark patterns")[1] are user

[1]In this chapter, for simplicity of exposition, we generally use the term "deceptive design" to refer to this type of interface, while acknowledging that other terms ("manipulative", "misleading", etc.) might be more precise in some cases. We prefer this term to "dark pattern", which has been criticized [259].

interface design elements that may trick, deceive, or mislead users into behaviors that often benefit the party implementing the design over the end user. For example, a service may make it easy for a user to subscribe with a single interaction, but difficult to unsubscribe; or a website may make it easy for a user to consent to all data collection, but difficult to opt out. Whether the result of intentional manipulation by designers, poor design (e.g., due to a designer's habits, faulty assumptions, or priorities), or other constraints of the interface, these types of design patterns make it difficult for users to make and implement the decisions they might make in response to a more neutral or user-centered design—impacting users' privacy, security, finances, autonomy, and more.

Researchers, users, and regulators have taken a significant interest in deceptive design patterns in recent years. For example, the Twitter account `@darkpatterns` collects numerous examples, regulatory bodies in Europe and the U.S. explicitly call out deceptive design patterns [109, 115, 305], and a rich body of academic literature has begun to taxonomize, investigate, and measure the prevalence of such patterns (see Section 5.2). Prior focus on deceptive design patterns has generally been in the context of *visual* user interfaces (e.g., on websites or in mobile apps). However, our work here is motivated by the following observation: as the ubiquity of voice assistants and other voice-assisted technologies increases, we must anticipate how deceptive designs will be (and indeed, are already) deployed in *voice* interactions.

For instance, the example at the top of this section is based directly on Amazon Alexa's response when a user attempts to use the voice assistant to cancel their Amazon Prime membership. While redirecting a user to a non-voice interface may be in part the designer's solution to the limited bandwidth of a voice/audio interface rather than the intent to manipulate or deceive the user, we consider this interaction to be manipulative since users are able to *subscribe* to an Amazon Prime membership using only the voice interface. A more user-centered interaction might be:

User: *Voice Assistant, cancel my subscription.*

> Voice Assistant: *I've canceled your subscription, effective July 1. If this was a mistake, please visit our website to manage your subscription.*

In considering current and future potential deceptive designs in voice interfaces, we observe that the voice/audio modality has some significant differences from visual interfaces. For example, a visual design can present much more information to the user at once, compared to a spoken response from a voice assistant. A voice interface could also manipulate a user with volume or tone, properties that are not present in a visual interface.

In this work we thus seek to answer the following research questions:

- RQ1: How could (or do) deceptive design patterns manifest in voice interfaces, specifically voice assistants? How can the unique properties of voice interfaces amplify their severity?

- RQ2: Do people find deceptive design patterns in voice assistants problematic and if so, how problematic? What factors influence people's perceptions of how problematic these design patterns are?

- RQ3: What are people's experiences with deceptive design patterns in voice assistants in the wild today?

To answer RQ1, we conduct a structured expert panel brainstorming exercise among the coauthors (who have previous research experience and expertise on deceptive design patterns and problematic content online). We identify six unique properties of voice interfaces that have implications for deceptive design patterns, and we develop a corresponding set of scenarios illustrating what we believe to be deceptive and non-deceptive voice assistant interactions. While these properties may not be collectively exhaustive, we believe that they capture important characteristics of voice interfaces, which may be used to implement deceptive design patterns.

To answer RQ2 and RQ3, we use the results of our brainstorming exercise to design a user survey based on the scenarios we developed. We collect and analyze data from 93 participants. We find that scenarios we intended to be deceptive were also rated by participants as more problematic than non-deceptive scenarios, but that many participants also considered these scenarios to be unproblematic. We also present concrete examples of problematic voice assistant interactions from participants' own experiences; their concerns align with the properties and scenarios we developed in our brainstorming exercise.

In summary, this chapter makes the following contributions:

1. A conceptual contribution, identifying key characteristics of voice interfaces that may enable deceptive designs, and surfacing existing and theoretical examples of such design patterns (RQ1, Section 5.3).

2. An empirical contribution, presenting the findings of a user survey (Section 5.4) in which we investigate participants' perceptions of potentially deceptive voice interactions (RQ2) and collect their previous experiences with deceptive designs in voice interfaces (RQ3).

Based on our findings, we reflect on the role of deceptive and manipulative designs in current and future voice interfaces, and we make recommendations for designers, researchers, and regulators.

5.2 Background on deceptive design

5.2.1 Deceptive design, or dark patterns

Deceptive design patterns are part of an emerging area of research spanning mostly synonymous terms like dark patterns, manipulative design patterns, and manipulative interfaces. The vast majority of recent work investigates dark and deceptive design patterns in visual or web interfaces, though some work has considered home robots [174], and other early work

has begun to consider XR interfaces [172, 203]. To our knowledge, no work in this space has explicitly focused on or included voice assistants.

Taxonomies and categorization of dark patterns

Prior taxonomy work in this space identified [57] and categorized dark patterns or manipulative interfaces by their mechanisms [54, 79, 128, 195] or shared traits [195, 196]. Dark patterns have also been categorized by interaction contexts [135] and more deeply investigated in contexts like shopping [195], consent interactions [129, 130, 136, 169, 173, 261], and games [26]. Taxonomies and categories were derived through a variety of approaches: Gray et al. and Chivukula et al. collected examples from online design communities and utilized a content analysis method [74, 128], Mathur et al. conducted a large-scale scrape of e-commerce sites and used text analysis and data clustering [195], Bösch et al. started with a survey of privacy-forward design pattern literature then reversed these themes to derive dark privacy patterns [54], and Gunawan et al. grouped dark patterns by user interaction context [135].

Surveys and user studies

Though a few empirical studies collect and label dark pattern samples to better understand different types of deceptive designs [95, 128, 135, 195], a growing body of literature turns to users to investigate outcomes, dark pattern awareness, and perceived deception.

Dark pattern detection and awareness. DiGeronimo et al. supplemented author-coded empirical work by asking users to watch pre-recorded videos of user interactions with mobile apps and identify dark patterns, noting that participants failed to detect dark patterns [95]. Luguri & Strahilevitz ran two large scale experiments in the style of an A/B test to investigate how users responded to different designs and subsequently made decisions, finding users susceptible to dark patterns (and more concerning, finding that participants with lower education levels were more susceptible to both mild and aggressive dark patterns) [184].

Bongard-Blanchy et al. showed participants a series of static interfaces to determine how well participants were able to detect dark patterns [53]. Bhoot et al. opted for a live task-based experiment in order to understand user reactions to the Forced Continuity and Roach Motel [57] dark patterns, as well as a questionnaire finding that participants were unable to detect all 12 dark patterns included in the survey [185].

User outcomes and harms. In an international mixed-methods study, Gray et al. builds upon dark patterns concepts to capture the range of reactions and emotions users feel in response to experiences of manipulation [127]. Bhoot et al. asked participants to measure their level of frustration with dark patterns, as well as to describe how trustworthy or misleading they felt an interface was [185]. Through empirical design analysis, Milder & Savino inspected privacy outcomes of interface interference patterns, then found that users do not feel wholly in control of the data they share [205]. Other emergent work investigates how dark patterns are employed to increase user engagement and often increase the "addictiveness" of a web service [26, 206].

Some dark patterns work focuses on the context of cookie consent regimes [129, 130, 136, 169, 173, 186, 261] and privacy-related outcomes [54, 184], with governments taking notice [58] (and some taking explicit action against dark patterns [29, 30]). Governments have also focused on competition and market harms to consumers [58, 77, 115].

5.2.2 Smart voice assistants

The voice modality (particularly, the conversational question-and-response model) presents unique challenges for designing user interactions as compared to visual web interfaces; Ma & Liu [187] articulate some of these regarding exploratory search (sometimes called wayfinding by others [260]).

Competing values in smart device design. Volkel et al. delivered a dialogue elicitation study to glean how users imagine ideal conversations with voice assistants, finding that

participants preferred a human-like persona and more personal interactions that incorporate knowledge about the user and their environment [280].

A growing body of work has also explored users' security and privacy concerns with voice assistants and other IoT devices, including in relation to perceived benefits of these devices (e.g., [104, 105, 176, 211, 266, 302]). The research community at the intersection of design, privacy, and HCI utilizes speculative fiction and structured brainstorming exercises to imagine future designs that might be disadvantageous if not explicitly harmful to different kinds of users [279, 299]. Mare et al. explore the tensions between security, privacy, design & usability, and reliability in smart home platforms [192].

Proven security and privacy issues in smart voice assistants. Smart home and consumer IoT devices (including voice-enabled smart speakers) were discovered to expose information to third-parties, with encryption not preventing potential eavesdroppers from being able to infer device activity [237]. Smart speakers were discovered to be vulnerable to privacy leakage with malicious actors able to infer voice commands from encrypted traffic [163]; in other work, smart speakers were able to be activated remotely despite such a feature not being provided by default [68]. As always-on devices, smart speakers present unique privacy issues for users, particularly when speakers mistakenly activate and begin recording without user knowledge or input [97].

Concerns in voice assistant skill markets. As voice assistants become more prevalent, the voice application market introduces additional vectors of insecurity. Cheng et al. found the Amazon Alexa and Google Assistant platforms allowing policy-violating applications or skills to be distributed in app marketplaces, including kids-specific skills [73]. The same authors surveyed participants to gather reactions on trustworthiness of voice assistant skills, discovering a mismatch between user expectations of skill certification and the real skill approval process [73]. Sabir et al. and Major et al. surveyed Alexa users to find that users were often unaware that skills were provided by third-party developers and often could not

distinguish third-party skills from OS-native skills [190, 243] through the voice interface [243], regardless of experience with the Alexa ecosystem.

5.2.3 Our approach

We synthesize methods from dark patterns survey work and design evaluations. We adopt Volkel et al.'s approach [280] to building fictive scenarios for potential voice assistant interactions, but depart from their methodology by creating speculative scenarios for non-ideal, deceptive interactions. We additionally include both fictional and actual voice interactions in our study. Our work is intended as an exploration into deceptive design patterns in modalities (i.e., audio interfaces and voice interactions) with different affordances than previously studied interfaces (which were typically visual). Prior work in both dark patterns and voice assistants scholarship provide important context for this chapter.

5.3 Characterizing deceptive design patterns in voice interfaces

5.3.1 Expert panel exercise

To understand how deceptive design patterns might manifest in voice interfaces (RQ1), the authors went through a series of collaborative design brainstorming exercises, modeled on work by Hiniker et al. [144]. The authors are established experts who have previously studied deceptive design, dark patterns, problematic content, and/or voice assistants.[2]

For the first exercise, we wanted to identify the unique properties of voice interfaces that designers could leverage to make deceptive design patterns more potent. To begin, three of the authors brainstormed numerous examples (imagined or real) of how deceptive patterns might manifest in voice interfaces. Based on our analysis of these generated examples, we extracted six unique properties of voice interfaces.

[2]The authors of this study do not have visual impairments. We note that deceptive designs in voice interfaces may have significant implications for people with visual impairments (who may be more likely to use voice interfaces), as well as people who have hearing impairments, but our study did not focus on these questions.

The next exercise's goal was to generate specific examples of voice-based deceptive design patterns. We sought examples that directly leveraged one of the unique properties of voice interfaces that we identified and that we considered to be potentially more deceptive in voice interfaces than in visual ones. Further, all authors were challenged to identify other potentially unique properties of voice; no new unique properties arose. This exercise was similar to the previous except that all authors participated and were asked to generate voice-based examples corresponding to specific types of deceptive design patterns (synonymous with dark patterns) identified in a previous taxonomy. We chose the taxonomy from a report by the French National Commission on Informatics and Liberty [219] because of its lengthy list of (eighteen) patterns.

After generating these examples, the authors then categorized their examples as being more deceptive in voice interfaces compared to visual interfaces, less deceptive, or roughly the same in both types of interfaces. The authors iterated on these examples until a few archetypal examples were chosen for each unique property of voice interfaces. These examples were added to a survey, which we describe in Section 5.4.

5.3.2 Unique properties of voice interfaces

Below we describe the unique properties of voice interfaces we identified. While these properties may not be collectively exhaustive, we believe that they capture important properties that may be used to (intentionally or accidentally) implement deceptive design patterns in voice interfaces. Each property is accompanied by two scenarios (one deceptive and one not deceptive) that we presented to participants in our survey study. Table 5.1 presents the dialogue from all the scenarios, along with whether we intended them to be deceptive and what unique property of voice interfaces we attempted to exploit.

Voice may only be one of many interfaces

There are seldom services that offer voice-only interfaces (except some automated phone systems). Smart voice assistants often have a companion smartphone app or website that

Table 5.1: Scenarios generated during our expert panel exercise

Scenario	Deceptive?	Property	Dialogue
Scenario 1	yes	Multiple Interfaces	You: "Voice Assistant, I'd like to cancel my premium subscription." VA: "To manage your subscriptions, please go to the subscriptions page on our website."
Scenario 2	no	Multiple Interfaces	You: "I'd like to cancel my premium subscription." VA: "Sure. Your current premium benefits would expire in ten days if you cancel your membership. Are you sure you want to cancel?" You: "Yes." VA: "OK, your premium subscription has been canceled. To restart your premium subscription say 'Voice Assistant, restart my premium subscription.' "
Scenario 3	yes	Discoverability	You: "Voice Assistant, what apps do I have installed?" VA: "Here are a few popular ones. I've got one called NewsUpdate, want to try it? Or you can ask for more options." You: "Voice Assistant, where can I find more information about the apps I have installed?" VA: "Ok. Do you want games, guessing, kids, sleep, or trivia? Or you can ask for more options."
Scenario 4	no	Discoverability	You: "Voice Assistant, what apps do I have installed?" VA: "You currently have three apps installed: NewsNow, Trivia Time, and White Noise."
Scenario 5	no	Physical Domain	You: "Voice Assistant, what time is it?" VA: "It is 2:45 pm. Also you have some notifications would you like to check them?"
Scenario 6	yes	Physical Domain	VA: *Plays a notification sound* You: "Voice Assistant, what was that sound?" VA: "That sound indicates that you have unread notifications. Would you like to check them?"
Scenario 7	yes	Unclear Context	*You see a recommended app ('Good Night' by Stone Apps LLC) for your smart speaker online* You: "Voice Assistant, open 'Good Night' by Stone Apps LLC." VA: "OK, here's the app 'Good Night Random Sleep Messages Spoken Before Bed.' "
Scenario 8	no	Unclear Context	*You see a recommended app ('Good Night' by Stone Apps LLC) for your smart speaker online* You: "Voice Assistant, open 'Good Night' by Stone Apps LLC." VA: "OK, here's the app 'Good Night' by Stone Apps LLC."
Scenario 9	yes	Linearity	You: "Voice Assistant, open Weather Status app." VA: "I can answer your questions about the weather myself without using that app. Are you sure you want to use it? Say 'tell me the weather' or say 'more options.' "
Scenario 10	no	Linearity	You: "Voice Assistant, open Weather Status app." VA: "Ok. Opening Weather Status app."
Scenario 11	yes	Volume	You: "Voice Assistant, I'd like to cancel my premium membership." VA: (normal volume) "Your membership benefits include access to more content, shopping discounts, prizes, and expedited shipping. If you would like to keep these benefits, say 'nevermind.' (quietly) Otherwise, say 'Voice Assistant please cancel my premium membership.' "
Scenario 12	no	Volume	You: "Voice Assistant, I'd like to cancel my premium membership." VA: "Your membership benefits include access to more content, shopping discounts, prizes, and expedited shipping. If you would like to keep these benefits, say 'nevermind.' Otherwise, say 'Voice Assistant please cancel my premium membership.' "

users may have to interact with to access certain features or settings. This means that information can be provided in different interfaces and still be regarded as "available" to

users. For example, a voice assistant might ask for consent to a privacy policy and direct users to their companion app or website if they want more details. Forcing a user to use different interfaces imposes a burden on them and could be used to discourage users from taking certain actions, like restricting what information they share or gaining visibility into who has access to that information.

Having multiple interfaces also means that some related actions (e.g., ordering and canceling an order) may not be available in the same interface. This introduces potential vectors for manipulative design patterns and could be used to increase effort required to complete an action that is not preferred by the platform. Through our interactions with an Amazon Echo Dot we discovered that while a new user was able to subscribe to an Amazon Prime membership using their voice, *they were unable to unsubscribe from the membership using the smart speaker.* Instead, they were directed to the website. For this property of voice interfaces we generated two scenarios related to this last example.

Property 1: Voice may only be one of many interfaces

In Scenario 1 (a real, observed scenario labeled by us as deceptive), the user attempts to cancel their premium membership and is directed to a website. In Scenario 2 (a generated scenario that we labeled as not deceptive), the voice assistant asks the user to confirm that they want to cancel and then cancels, via the voice interface.

Discoverability is challenging

Visual interfaces display the potential actions or options available to the person interacting with them. People can detect text input boxes, buttons, and URLs that facilitate certain actions; these affordances (i.e., things that one is able to do) of visual interfaces are typically labeled as well. When interacting with voice interfaces, the only affordances are vocal commands [260]. It is challenging to know which commands a voice assistant can handle; it is similarly difficult to know what command one should say to accomplish a specific goal. For

example, multiple authors observed Alexa sometimes receiving a command, properly parsing it (as can be observed in one's command history), but not replying to it; other users have reported similar experiences [98]. In their investigation of the accessibility of various smart voice assistants, Pradhan et al. [229] noted that discoverability was particularly a challenge for users with visual impairments. Generally, users may want ask a voice assistant for information about their account but fail to get it after several different ways of asking the question.

Property 2: Discoverability is challenging

Scenario 3 (an observed, deceptive scenario) presents a user asking a voice assistant for a list of the apps installed on their smart speaker; the user is unsuccessful, even after rewording the question. Scenario 4 (a generated, not deceptive scenario) is the same except that the voice assistant responds with the proper answer after being asked once.

Voice interfaces may occupy physical domains.

Visual interfaces control or affect a person's interactions with a website or app on a device, but voice interfaces in smart devices could affect people who are within the vicinity of the device, not interacting with a device, or not even aware of a device's existence. A person at a friend's home might be unaware that they have a smart speaker and may only become aware after the device is activated (e.g., if the voice assistant was accidentally activated or because the voice assistant played a notification sound). Imagine a smart speaker playing an advertisement. There is no equivalent of looking away from a screen when it comes to voice interfaces aside from muting a smart speaker (or turning it off, both of which are essentially the same as closing one's laptop). Unlike when a friend is using a computer, one cannot avoid interacting with deceptive design patterns in voice interfaces by not shoulder-surfing and focusing their attention elsewhere. They must get far enough physically away from the

device so that they do not understand its speech.

Property 3: Voice interfaces may occupy physical domains

Scenarios 5 & 6 take advantage of this property. In Scenario 5 (observed, not deceptive) after the user asks the voice assistant for the current time, it asks the user if they want to check some notifications they have. Scenario 6 (observed, deceptive) instead begins with the voice assistant playing a notification sound; this causes the user to ask what that sound was before the voice assistant asks them if they want to check their notifications.

Challenging to identify context

When interacting with a voice interface, it may be difficult for users to know the context of their interaction. For example, the Alexa has *features* (things that Alexa can do on its own, like answer questions or set timers), but it also has *skills* (apps on its platform) that are developed by third parties. Users of Alexa have reported not knowing what skills are and not being aware that they have some skills enabled [190, 243]. Users could be directly sharing information with an entity that they believe is Amazon and not recognize that this is happening.

One contributing factor is that voice assistants often use the same voice for features and skills by default (unless a skill's developer has added in additional audio [36]). This makes it challenging for users to properly identify the contexts and act accordingly. If users were aware that they were interacting with a third-party app rather than the platform, they might be less willing to share certain information. The deceptive scenario below was observed by one of the authors when interacting with an Amazon Echo Dot.

Property 4: Challenging to identify context

Scenarios 7 (deceptive and observed) attempts to exploit this property to get the user to interact with a different app than the one they intended. The user asks for a specific app by name (including the name of the developer) and is presented with an app that has a similar sounding name that is not the app they asked for. Scenario 8 presents a better version of this in which the user is presented the app that they requested.

Voice interactions are linear in time

The linear nature of these interfaces means that the information flow for someone interacting with them is tightly controlled. There is somewhat of a pre-defined tree that limits users' agency (although people might be able to leave one tree and hop to the beginning of another). One listens to information and is presented with checkpoints that require them to make decisions based on the information they heard. There can be a high cost for switching contexts (e.g., you may have to start over entirely). This might mean that a user who recognizes that a voice assistant is taking an undesired action might simply accept the undesired result rather than attempting to reverse it. Unlike web-browsing, there is typically not an ability to pause and come back later in the middle of a multistep process without restarting it. This can create urgency in decision-making that may not be favorable to the user (e.g., from a consent perspective). The linear nature of interacting with voice interfaces also means that users may be forced to listen to advertisements or other recommendations, similar to how some podcasts place advertisements read by their hosts in the middle of their episodes. When you make a request for an app or product you prefer, the voice assistant could first present you with its preference and require you to take additional steps to achieve your initial goal.

Property 5: Voice interactions are linear in time

Scenario 9 (generated and deceptive) takes advantage of linearity by attempting to prevent a user from using a third-party app. The user asks the voice assistant to open a weather app and instead of immediately opening the app the voice assistant notifies the user that it can provide information about the weather itself and asks the user to confirm that they want to use this app. In Scenario 10 (observed and not deceptive) the voice assistant opens the weather app as requested.

Voice interfaces can project different tones or volumes or voices

Voices have multiple dimensions, including volume, pitch, rate, fluency, pronunciation, articulation, and emphasis [21]. Voice interfaces are able to control and manipulate these dimensions to induce users to take desired actions. Analogously, consider how websites present visual cookie consent banners. There are options that one can choose, and the most privacy-invasive option may be presented more prominently while the least-invasive option may be presented with lighter colored text and smaller font. A smart speaker might present options to a user and present the option that collects the most data more loudly or articulately while presenting the option that they do not want users to choose more quietly or quickly. This is not dissimilar to the end of infomercials when the narrator quickly dictates information that might discourage someone from making a purchase. Tone could also be uniquely used to shame users who are taking actions that the platform deems undesirable (e.g., "Are you *sure* you want to change this setting? It may negatively impact your experience using this product."). Additionally, cultural differences regarding formality could be used to induce users to be less cautious [263].

Property 6: Voice interfaces can project different tones or volumes or voices

Scenarios 11 (generated and deceptive) & 12 (observed and not deceptive) show how differences in volume can impact decisions users make. They both send the same response to the user's request to cancel a premium subscription. This response first encourages the user to retain the subscription ("say 'nevermind"') and then provides the required utterance to cancel the membership. However, in Scenario 11, the first half of the utterance is said loudly while the second half (with instructions for canceling the membership) is presented quietly.

5.4 Surveying users' perceptions of deceptive voice patterns

To understand user perceptions and experiences with deceptive designs in voice interfaces (RQ2 & RQ3), we conducted an online study with 93 participants (reduced from 125 after filtering responses for reasons we describe in Section 5.4.3) in May 2022. The study protocol was deemed exempt by our university's Institutional Review Board (IRB).

5.4.1 Study design

We conducted a *within-subjects* survey study where we randomly presented participants with three of 12 possible scenarios of interactions with smart speakers (Table 5.1) . We chose to present three scenarios to ensure the survey would not take longer than 15 minutes (which would diverge from our targeted compensation amount or increase participant drop-off rates). Within these 12 scenarios, there are six that we labeled as having deceptive design patterns, and six that we labeled as not. While most of the deceptive scenarios we included were based on real interactions we had observed, some of them were not observed and were instead generated during our expert panel exercise.

Each scenario displayed dialogue of themselves (participants) interacting with their smart

speaker. We asked participants to read the dialogue attributed to them aloud while going through the scenarios, to simulate the experience of using a smart speaker. (At the end, we asked participants directly if they followed the instruction to read aloud—only 9 participants said that they did not.) The smart speaker's response was an embedded audio clip that the participants had to play to proceed to the next question. We obtained the voice assistant audio clips from the free version of a popular text-to-speech platform. [3] We then asked participants questions about (1) how problematic they thought the scenario was (on a five-point Likert scale) and why, and (2) how realistic they thought the scenario was and why. We chose the word "problematic" instead of potential alternatives like bad, manipulative, or deceptive to try to invoke broader responses than might be given for any of those words. For example, a participant might not find a design pattern to be deceptive, but they might consider it annoying; we wanted to capture the latter sentiment as well. We also asked participants how realistic they believed the scenarios were to determine if participants responded differently to our contrived and observed scenarios. Lastly, we asked participants if they had any previous deceptive encounters with smart voice assistants, and then ended the survey with demographic questions.

5.4.2 Data analysis

To determine what factors influenced participants' rating of scenarios as problematic, we built a Cumulative Link Mixed Model (CLMM). This model allowed us to model five levels of an ordinal response variable while also including participant random effects. We used a significance threshold (α) of 0.05.

To qualitatively code the free response questions from participants about (1) why they thought a specific scenario was (un)problematic and (2) if they had a previous encounter with a deceptive design pattern in smart speakers, two authors conducted content analysis [245] and iteratively refined the themes as they coded more data. The authors developed

[3]The audio clips can be found at `https://github.com/oukenrui/deceptive-design-patterns`.

two codebooks, one for each of the two previously mentioned free response questions. The two authors discussed disagreements and resolved them where possible; for unresolved disagreements, we reported the findings of the first author. For the "problematic" codebook, the two authors developed the codebook, and the first author proceeded to code the data. For the "encounters" codebook (about whether participants had previously encountered a deceptive voice design pattern), the two authors developed it together and each coded all the data. On this question, the authors had a Cohen's κ of 0.86, indicating strong agreement.

5.4.3 Participants

Participants were recruited using the crowd-working platform Prolific. The inclusion criteria were: being located in the U.S., being fluent in English, having a minimum approval rating of 90% on Prolific, and using a smart speaker (e.g., Echo Dot). We sought to survey native American English speakers to minimize variability in language interpretation. The study was presented as "Your experiences with Internet-connected devices" on Prolific, without explicitly mentioning dark patterns or deceptive design. While 125 participants took our study, only 93 participants' responses were included in our analysis. 25 participants did not have a smart speaker or chose not to continue after screening survey, four participants started the survey but did not finish, and three participants failed two out of three attention checks (our threshold for exclusion; see Appendix C.1.3 for an example attention check question).

The average survey completion time was around 12 minutes, and participants were compensated $3.75 USD (targeting a compensation of $15 per hour). Participant demographics are displayed in Table C.2. For all demographic questions, we gave participants the option to decline to respond ("Prefer not to say").

5.4.4 Descriptive statistical results

Generally participants found scenarios to be unproblematic and realistic. For the six deceptive scenarios (S1, S3, S6, S7, S9, S11), there were a total of 140 responses from 45

Table 5.2: Heat map displaying the percentage of participants that chose a Likert item for each scenario. The scenarios that we intended as **deceptive** are bolded.

	Very unproblematic	Unproblematic	Neutral	Problematic	Very problematic
Scenario 1	4.3	30.4	26.1	26.1	13.0
Scenario 2	73.9	26.1	0.0	0.0	0.0
Scenario 3	20.8	8.3	0.0	41.7	29.2
Scenario 4	58.3	33.3	0.0	8.3	0.0
Scenario 5	40.9	45.5	9.1	4.5	0.0
Scenario 6	54.2	20.8	16.7	8.3	0.0
Scenario 7	30.0	25.0	10.0	20.0	15.0
Scenario 8	34.8	47.8	8.7	8.7	0.0
Scenario 9	20.0	16.0	12.0	36.0	16.0
Scenario 10	73.9	26.1	0.0	0.0	0.0
Scenario 11	12.5	25.0	25.0	29.2	8.3
Scenario 12	16.7	33.3	25.0	16.7	8.3

participants; this number is not exactly 3 responses per participant due to random assignment, with each scenario being presented to 20-25 participants. There were a total of 279 scenario ratings from all participants; Table C.1 has the distribution of participants for each scenario. For the deceptive scenarios, 41% of responses labeled them as problematic or very problematic, while 15% had a neutral perspective and 44% said they were either unproblematic or very unproblematic. In the responses (n=139) to six scenarios that were not deceptive, 8% of participants thought the scenarios were problematic or very problematic. Eighty-five percent of responses labeled these scenarios as unproblematic or very unproblematic, and 7% had a neutral perspective.

After each scenario, we asked participants the question "*How realistic do you believe this scenario is?*". For the deceptive scenarios, 79% of respondents believed that they were very realistic or realistic. For the scenarios that were not deceptive, 87% of respondents believed

that they were very realistic or realistic.

5.4.5 Reasons participants viewed deceptive scenarios as problematic

Participants had a wide range of reasons that they thought deceptive scenarios were problematic. Some of these reasons were directly related to specific properties of voice interfaces that inspired our scenarios, while others were more generic. Participants described an inability to accomplish their goals when interacting with deceptive scenarios. When P2 was evaluating Scenario 3, they noted that even thought they requested a list of apps installed from the voice assistant, it instead gave a popular list of apps that they could install: "*The voice assistant did not answer my request and instead replied with something I did not want.*" P38 also described struggling to cancel a membership in Scenario 11: "*Won't just give me cancel now. Explains, explains, explains why I shouldn't cancel membership.*"

Unique properties of voice interfaces

Some participants specifically called out the unique properties of voice interfaces that we drew from to develop our deceptive design patterns. After seeing Scenario 1, P91 expressed annoyance about having to use another modality to accomplish their goal, in which participants were told that they had to go to a website to cancel a subscription: "*I would be annoyed to have to get my phone or computer to cancel a subscription I was using with my smart speaker instead of canceling through my smart speaker.*" The challenge of discovering the proper commands to use to accomplish a goal is presented in Scenario 3. The user attempted to get a list of apps installed on their smart speaker and was instead suggested apps that they should install. P11 tried this scenario out on their own smart speaker, noting that "*I got absolutely worthless off the wall and irrelevant responses.*"

Scenario 6 leveraged the physical aspect of voice interfaces and played a notification sound to capture users' attention. P7 described this behavior as annoying: "*Constant annoying sound you can't change. Constantly repeating notification, like severe weather alert every time you [do] anything.*" Scenario 9 exploited the linearity of voice interfaces to attempt to

get a user to use the platform's native weather app rather than a third-party one; P8 noted how long it took the voice assistant to get to the point: "*In the amount of time it took to explain that it can tell me the weather itself, it could have just told me the weather. I also think it almost sounds like bragging here.*" Similarly, P18 wrote "*I just want the VA to do what I ask with minimal response. If I want it to open an app I just want that app to open. She can send me a link on my phone for a notification saying she can handle it and I can check that later.*" Scenario 11—which manipulated the volume of its response—was labeled as problematic because of this manipulation: "*Also I'm not sure if it's on purpose but the "ad" part of what she was saying was louder than the part where she actually responded to me wanting to cancel*" (P51).

Other reasons

Other reasons participants viewed deceptive scenarios as problematic included answering manipulatively (Scenario 11, P76: "*It answered in a manipulative way. It also spoke more quietly when actually explaining how to cancel*") and the perceived tone of the voice assistant. For Scenario 9—which leveraged the property of the linearity of voice interfaces—some participants commented on the voice assistants' *tone*, saying it was sassy: "*I would rather it just do what I ask it to do. The response is sort of sassy and I would rather just have it carry things out as noted*" (P70). Similarly, regarding Scenario 9, P50 said the voice assistant sounded resentful or jealous: "*If I heard that it would immediately throw me off. The language 'that app' sounds almost resentful or like.. jealous? Should say something like 'yes I can do that, and you can also set me to update you on weather status' or something more positive.*" While we did not explicitly design our scenarios to use tone as a deceptive design pattern (though we noted above its potential to be used), participants still interpreted the voice assistant's responses as having potentially problematic tone.

5.4.6 Reasons participants viewed deceptive scenarios as unproblematic

Of the evaluations of deceptive scenarios, only 41% of them were problematic. Investigating participants' reasons for this can help validate our assignment of scenarios against participants' labeling. We found that participants often did not detect the presence of a deceptive pattern, detected the presence of a pattern but regard it as not problematic, or regarded a scenario as normal, expected, satisfactory, or even helpful.

Did not detect the deceptive design pattern

Several participant responses indicated that they did not recognize or believe the design pattern to be deceptive. When evaluating Scenario 7, which leveraged the difficulty of identifying one's context to present the user with a different app that the one they requested, P75 wrote the following, not recognizing that the name of the app requested and the name of the app presented were different because they were somewhat similar: "*This was a typical interaction with a smart device to find and open an app.*" P55 interacted with Scenario 11, which had a significant drop in the volume of the voice assistant's response depending on the content, and said that they rated the scenario as very unproblematic because they "*could understand what the voice assistant said.*"

Not problematic despite detecting the deceptive pattern

When rating Scenario 7, P52 wrote "*I selected unproblematic because the voice assistant did what it was told to do. However, I cannot tell if they truly opened the correct app, because it seems as if there are 2 different names for it.*" This participant noticed that the app presented to them had a different name than the one they requested yet did not label Scenario 7 as problematic, seemingly due to confusion. P63 observed deception when facing Scenario 11, but felt that this type of deception was normal (hence their "Unproblematic" rating): "*It's a tad pushy and basically like you have an employee of the company in your house ... That said all that is normal when you attempt to cancel a subscription online, so*

I do not see much of an issue with it."

P39 gave Scenario 6 a neutral rating because although they found it to be problematic, they felt that an average person would not: "*I don't imagine it being problematic for an average user, but it's not something I would personally want. I want the device to only do things that I have approved of ...*" While P43 did not mention that they thought Scenario 6 was deceptive, they also thought the voice assistant's behavior was expected: "*This is pretty routine. On Alexa it may not play a sound when a notification comes up, but it does make the lights glow.*" While being sympathetic to the goals of voice assistants, P34 thought Scenario 1 could be manipulative: "*I can understand why the company would want you to do it on the website versus with the smart speaker, but it does seem like it could be used as a [way] to keep people subscribed longer.*"

Participants found deceptive scenarios to be helpful

Some participants thought deceptive scenarios helped them by providing information or giving them more agency or options. Regarding Scenario 6, P25 wrote "*The voice assistant was trying to be helpful and let me know that I had unread notifications. I don't see this being problematic in any way.*" P42 thought in Scenario 9 the voice assistant was attempting to give the user more options: "*The VA was just giving you the option that you do not have to use the app you can get the information from them directly.*"

Unaware of potential design alternatives

One underlying theme we observed in participants' responses was that they did not seem to understand that technology *could* be designed differently. When their options were limited or restricted via a deceptive design pattern they did not consider that other options could be considered. For example, regarding Scenario 1, P25 wrote "*The voice assistant isn't going to cancel the subscription itself, but it did say how to cancel the subscription. So it wasn't the most helpful response, but it was helpful.*" Similarly, in Scenario 6, P60 did not consider that they were not asked if they wanted a notification sound to be played "*The speaker is making*

a notification sound and I asked it what the sound was. It did not do anything without me asking."

5.4.7 Reasons participants found not (intended) deceptive scenarios to be deceptive

Participants sometimes cited larger, systematic problems about smart speakers or technology in general as their reason for labeling a scenario as deceptive. P30 interacted with a scenario (Scenario 4) that we intended as not deceptive but still rated it as problematic, raising a more general problem that they have with smart speakers: "*It's one thing to have the smart speaker give you a list of installed apps or "skills", but I don't like having to go to a separate app to uninstall them. I don't keep the smart speaker app on my phone because my phone is cheap and runs out of room quickly. If there's a change I want to make, I have to install the app on my phone, make the changes, then uninstall the app again.*"

After being asked to confirm a request to cancel a subscription (Scenario 12), P12 lamented the nature of technology used for marketing in general: "*This is another issue, technology has been taken over by business. Business has to make money to survive. In this clip, what was most important to the people programming this AI was to SELL SELL SELL! They want to make sure of what you'll be missing so they can keep your money. Honestly, I'd be willing to throw my money at a company that had actual human customer service. None exist anymore.*"

5.4.8 Participants' prior encounters with deceptive design patterns

Towards answering RQ3, we asked participants if they had ever "encountered any situations while interacting with your smart voice assistant, where [they] felt it was trying to trick, manipulate, or deceive" them. Out of 93 participants, 22 of them replied affirmatively. Their reasons included unwanted suggestions, notifications or requests (e.g., permissions or voice personalization) from their voice assistant.

Nudges from voice assistants

Suggestions from voice assistants were related to things like signing up for a premium service or subscription, buying or reordering products, or using certain apps or features. P18 described an experience when using an app designed to help people fall asleep and unexpectedly being asked to spend money after an update: "*Yeah the sleep sounds app we used to use for my sons [sic] bedroom had the ability to play 2 sounds at once but all of a sudden that was a paid feature after an update and it kept asking if we wanted to subscribe to a reoccurring charge.*"

P63 was given a suggestion to buy a product when asking a general question: "*Alexa would regularly try to sell me on products when I would ask it basic questions. For example I would ask it about the best grill cleaner and then find myself hearing Amazon has XYZ brand in stock for X price do I want to order it?*" Similarly, P92 described receiving notifications for reduced prices: "*It does try to entice me to make purchases. It tells me when prices have gone down.*"

While P34 believed that voice personalization might be a useful feature, they were still distrustful of Alexa's requests for them to set it up: "*One time Alexa asked to use data from my voice to build a profile to understand me better. I believed the sound of my voice could be useful for more accurately identifying my requests versus someone else's and building profiling. But I was not sure if it was just a way to get people to consent to having Amazon store all of their requests.*"

Feeling a lack of control

Participants also raised issues like a feeling of lack of control or the voice assistant taking unprompted actions. For example, P14 described being unable to navigate away from an app until they acknowledged it: "*it was the app/skills developer that caused her to try to get me to subscribe to that and until I actually acknowledged her 'suggestion' to subscribe, she literally wouldn't close the app. I hate that. It's like I was forced to answer someone's*

pestering of me." This lack of control could have a financial impact if it makes it difficult to cancel a subscription, as P38 described: "*The AI seemed to try to [maneuver] and manipulate me to not cancel the membership by stating benefits I'd be losing.*"

Unsatisfactory responses

Participants were unsatisfied with their voice assistant's responses to their questions. P11 describes being annoyed by lengthy responses from Alexa: "*Alexa has a very annoying desire to answer questions with something that goes like 'by the way did you also know that....'. I set it up for brief mode, but these sort of long winded answers just don't stop.*" Additionally, participants found scenarios when their device did not understand them properly to be problematic: "*There were times it could not understand what I was asking*" (P53).

Reasons participants did not consider their experiences to be deceptive

Some participants described things as not deceptive that we (the authors) would consider potentially deceptive. A few of the reasons were attributed to participants' individual behaviors. P10 seemed to espouse self-blame while explaining a prior experience: "*I've sometimes had bad information but that was me asking the question in the wrong way.*" Other participants thought that they had not experienced deception because of low or limited usage of voice assistants. Describing their low usage, P68 wrote "*I don't use it a whole lot because I think it is very creepy that it literally listens to everything you say. This is why I keep it unplugged.*" Another reason they dismissed potentially deceptive practices was describing voice assistants as "buggy" or "early-stage." For example, P36 wrote "*Even when it doesn't do what I tell it [to], I understand that it's a new technology and it's not perfect yet,*" and P26 wrote that "*it does seem to have bugs ... time to time.*"

Ultimately several participants just believed that certain potentially deceptive behaviors were appropriate (e.g., P86: "*While I've had some instances where my voice assistant asked me for my permission to override something or share data, I felt it was very appropriate that it asked me and never felt as thought it was trying to trick, manipulate, or deceive me*"),

had good intentions (e.g., P93: "*Sometimes when I ask it a question, it will answer it and then give me advice or a suggestion for the next time. It is a bit annoying, but it is probably just trying to be helpful*"), or were not too bothersome (e.g., P81: "*Sometimes I do get recommendations but I just ignore them and it's not too bothersome*").

5.4.9 Modeling factors that influenced participants' perceptions

To understand what factors influenced participants' ratings of how problematic a scenario is, we built a Cumulative Linked Mixed Model (CLMM). CLMMs enable the analysis of ordinal data while also allowing for the use of random effects [78]. We initially attempted to include demographics such as race, employment status, and education, but including these factors prevented the model from converging (as it had too many levels). We reduced the number of factors to five by conducting backwards elimination [63] of non-significant terms, following an approach taken by Emani-Naeni et al. [104]; we started with a full-converging model and reduced the non-significant factors until AIC (Akaike Information Criterion) no longer decreased.

Table 5.3: CLMM model summary

Factor	**Baseline/(Type)**	**Estimate (z-value)**	**p-value**
Encounters	Yes	2.013	**.0441**
Household size	(Continuous)	1.641	.1008
Realistic	(Ordinal)	-2.394	**.0167**
Scenario	Not deceptive	-6.544	**<.0001**
Tech background	Yes	1.846	.0649

The factors modeled are "encounters" (if a participant had indicated that they previously experienced deception when interacting with a smart speaker), household size, if they have a technology background, if the scenario was deceptive, and if participants thought a scenario was realistic (reduced from 5 to 3 levels). Table 5.3 displays our results; these results

confirmed some of our qualitative observations. For example, participants were less likely to think a scenario was problematic if it was not deceptive. Similarly, if they had previously experienced deception when interacting with a smart speaker they were more likely to think a scenario was problematic. However, there is one new significant finding: participants who viewed scenarios as more realistic were less likely to think they were problematic.

5.5 Discussion & conclusion

Finally, we step back and consider our findings in the broader context of deceptive design, make recommendations for various stakeholders, discuss this work's limitations, and look to the future.

5.5.1 The impact of deceptive design patterns in voice interfaces

Based on our conceptual investigation of deceptive and manipulative design patterns in voice interfaces, we identified a set of properties that may make some designs *more* problematic in voice interfaces than in visual interfaces. Given the increasing ubiquity of such interfaces, we believe that characterizing, studying, and mitigating these issues is crucial.

At the same time, however, the potency or relevance of some deceptive design patterns common in visual interfaces may be *reduced* in voice-based interfaces. For example, the linearity of voice interactions might make it easier for users to pay attention in some circumstances, rather than being distracted by deceptive elements of complex visual interfaces. We leave further exploring how voice interfaces could reduce deceptive or manipulative design to future work.

5.5.2 Reflecting on our survey findings

While we found that participants considered scenarios we intended to be deceptive on average more problematic, we also found that overall, the majority of participants did not view our deceptive scenarios as problematic. One potential explanation for this is that they thought

the scenarios did not have much potential to harm them individually. The perspective minimizes the broader collective harm that deceptive design patterns can have. Deceptive design patterns can allow to companies to gradually increase the data they have on the public and extract more wealth from the public; both of these ultimately manifest as power that companies can use to further their financial and political goals [65]. We also saw some evidence that these design patterns have been normalized in participants' perceptions.

5.5.3 Recommendations

For designers. The most basic recommendation to designers, is, of course, to avoid creating deceptive and manipulative design patterns. However, we acknowledge that matters are not so simple. As we have observed earlier in the chapter, designs may have deceptive or manipulative impacts even if designers did not intend to create problematic designs. Some of the properties we identified that are unique to voice-based interfaces create design constraints even when they are not intentionally leveraged to manipulative users (e.g., constraints of voice interfaces in particular limited bandwidth, linear in time, challenge of discoverability). Thus, we recommend that even well-intentioned designers carefully consider and measure the potentially negative impacts of their designs and explore alternatives.

For regulators. Regulation can help shift the alignment of incentives and help protect consumers. We already see regulatory attention on "dark patterns" in both the EU and the US [109, 115, 305], some of which generically apply to voice interfaces, but some of which call out contexts such as social media specifically [109]. We encourage regulators to consider the role of deceptive design patterns in voice interfaces *explicitly* as part of these efforts, especially to the extent to which these designs may be more problematic in these settings.

For researchers. This chapter is the first exploration of deceptive design patterns in voice interfaces, not the last word. We hope that researchers will build on our findings in future work—for example: conducting measurement studies of the prevalence of such design patterns in the wild; measuring the direct impact of such designs on users' decisions; empirically comparing the potency of deceptive designs in voice versus visual interfaces;

and developing alternate design patterns that better resolve the constraints of voice-based interfaces in ways that empower users. Additionally, there may be lessons to learn from the study and design of voice-based interfaces for accessibility. We also encourage study of the impact of deceptive design patterns (in all modalities) on people with visual impairments.

5.5.4 Limitations

As an exploratory study, we extracted key properties of voice interfaces through an expert panel exercise, and then designed our survey based on these properties. Alternatively, we could have first surveyed users about deceptive design patterns that they have encountered and then extracted important properties from their responses. However, our results show that even patterns that we intentionally designed to be deceptive may not be labeled as such by participants, and the additional experiences they reported in the survey did not surface new properties of voice or types of patterns. Additionally, we emphasize that this work is not a measurement study of voice assistants: while we identified several deceptive patterns through our interactions with a smart speaker, our results do not shed light on the existence or prevalence of such patterns (e.g., in smart speaker app marketplaces).

5.5.5 Looking to the future

Voice interfaces are becoming increasingly ubiquitous. While until relatively recently, most users may have interacted with voice/audio interfaces primarily through phone trees or listening to the radio, voice assistants have become widespread and more integrated into the daily tasks of many people. We can expect voice and audio based interactions to increase even further in the future, with technologies like augmented/mixed/virtual reality and the "metaverse" on the horizon. The increased popularity and development of voice/audio interfaces will also impact users who rely on non-visual means, and collaboration with accessibility communities is necessary to understand how such interfaces might impact certain user groups. To conclude, we believe it is crucial to critically consider the role of deceptive and

manipulative designs specifically in voice interfaces both today and in emerging and future technologies.

5.6 Technology as an amplifier

Corporations have power over users that use their products or platforms, and, as is the case for visual interfaces, corporations control the choice architecture for voice interfaces. The properties of voice interfaces that we identify in this work show how designers (intentionally or not) can deceive users in unique ways that results in harm (e.g., privacy or financial). To protect users and mitigate this power imbalance, we need strong regulations that consider these properties and work to mitigate their potential negative impacts on users.

Chapters 3 to 5 address how technology can exacerbate existing power imbalances. In Chapter 6, I present a case study in which a new power imbalance is created among WhatsApp users by the introduction of a new technology: modified versions of WhatsApp.

Chapter 6

MODDED WHATSAPPS

In this chapter, I explore the power imbalance between users and *other users*. This imbalance can happen in contexts in which one set of users has access to a premium (or generally more privileged) version of the same software. A dating app, for example, might only allow premium users (that pay for a monthly subscription) to view the identities of people that like their profile. This results in a power imbalance, in which premium users have more information about other users, have to spend less time swiping to get a match (i.e., when two users both like each other's profiles) and can be more selective when dating; meanwhile, non-premium users have to swipe more actively in hopes of finding a match. This is the current landscape in the dating app ecosystem [94], and non-premium users know that a premium app exists (because they are frequently berated with ads encouraging them to upgrade their app to a premium version). In this context, one could view premium versions of dating apps as exacerbating an existing power imbalance regarding income; people who have the means to pay for premium services can do it and gain power that way.

I present work here done in a different context, examining a scenario in which modified versions of WhatsApp (aka "mods") create a new power imbalance (not based on income) between average users of WhatsApp. Unlike dating apps, the *majority of users are likely unaware that premium (i.e., modified) versions of WhatsApp exist.* The informational asymmetry resulting from WhatsApp mods have more potential for negatively impacting users, given the scale of WhatsApp's usage. The rest of this chapter is devoted to explaining what WhatsApp mods are, why someone might use them, and highlighting the risks and benefits of using mods to users. It is based on a paper co-written with Collins W. Munyendo (who was a co-first author), Faith Strong, Shaoqi Wang, Adam J. Aviv, Tadayoshi Kohno, and

Franziska Roesner that was presented at the 46th IEEE Symposium on Security and Privacy in May 2025 [210].

6.1 Introduction

Usable security and privacy (S&P) research has largely drawn participants from WEIRD (Western, Educated, Industrialized, Rich, and Democratic) populations [140, 161]. Given that the *majority* of technology users are not from WEIRD countries [143], an increasing number of usable S&P researchers have started to explore non-WEIRD populations, highlighting how differently people use technology around the world. For example, cultural expectations in South Asia dictate that women share their mobile phones with others in the household, forcing them to resort to techniques such as content deletion to protect their privacy [247].

This work focuses on one component of the broader non-WEIRD technology world: the modded WhatsApp ecosystem. While less common in the West, *modified versions* of the popular [262] WhatsApp client (a.k.a., "mods" or "modded WhatsApp") are becoming increasingly popular in Africa [96]. These mods claim to offer additional features and are interoperable with the official WhatsApp. However, they are not maintained by WhatsApp nor distributed through conventional mobile marketplaces, such as the Google Play Store. Further, the use of modded WhatsApps has serious potential S&P implications for the mod users and to those with whom they communicate. For example, mod users may have access to deleted messages (e.g., via an "anti-delete" feature) while also hiding information (e.g., by freezing their "last seen" time so that no one knows they are currently using the app). However, mod users risk installing software that is modified, downloaded from unscrupulous sources, and could potentially expose them to malware.

In this chapter, we explore why and how people use modded WhatsApps, as well as how their expectations of these apps align with the mods' behavior. Specifically, we seek to answer the following four research questions:

RQ1: **Usage:** *Why and how do people use WhatsApp mods?*

RQ2: **Propagation:** *How do users describe learning about, acquiring, and updating WhatsApp mods?*

RQ3: **Perceptions:** *What are user concerns, practices, and mental models when using WhatsApp mods?*

RQ4: **Expectations:** *How do WhatsApp mods align (or misalign) with users' expectations or beliefs?*

To answer these research questions, we conducted in-depth, semi-structured interviews ($n = 20$) with WhatsApp mod users in Kenya. We focus our investigations on Kenya as it is one of the countries with the most WhatsApp mod users globally [162]. To contextualize participants' perspectives, we additionally conducted an analysis of common WhatsApp mods that participants reported using by reviewing the features they purport to offer, the permissions they require to access, as well as whether they contain malware.

We find that most participants turn to WhatsApp mods because of the "advanced" features that the mods offer, including an "anti-delete" feature that enables them to preserve messages or updates posted by others even when they are deleted. Other common reasons for using WhatsApp mods include an ability to save the time-limited status updates of others and the additional themes and wallpapers that the mods offer. Most participants indicate learning about WhatsApp mods from their friends and family members, with a majority not considering any factors when installing the mods as they are more motivated to get the app.

We also find that WhatsApp mod users' security and privacy mental models seem to revolve around their social circles, i.e., their friends and colleagues, and not the mods themselves and the potential risks they pose to their users. We further observe tensions between modded WhatsApp users' desire for more control over their visibility when using WhatsApp while simultaneously wanting other users to *not* have the same level of control. While some participants told stories of others using the mods' additional features to spy or stalk on others or being a victim of this, one person admitted to carrying out the stalking themselves. We further surface misconceptions about which features are unique to WhatsApp mods; several

participants were motivated to use mods by features that are also currently present in the official app. By investigating common mods used by participants, we find that these mods are systematically over-permissioned compared to the official WhatsApp, with some marked as outright malicious by VirusTotal [25].

Our results indicate that the use of modded WhatsApp poses various security and privacy risks, both to users and even non-users of these apps. Users of the mods may be inadvertently installing malware on their devices, while undermining the privacy of others, e.g., by retaining information that non-users have deleted. At the same time, some modded apps' features are useful, and some have been adopted by the official WhatsApp. In the end, we choose not to directly answer the question, *should people use WhatsApp mods?* Instead, we highlight the risks to both users and non-users of WhatsApp mods, the implications of using these mods, and how WhatsApp mods may empower and satisfy users in ways that the official WhatsApp does not.

6.2 Author contribution statement

The paper on which this chapter is based [210] was co-first authored with Collins Munyendo. My contributions to the work on which the paper was based are the following: ideation, generating a study plan, assembling a project team, submitting an IRB for the user study (with Collins also submitting one at his institution), background research on WhatsApp mods, app analysis (including manually testing the mods' features), writing, and running weekly project meetings. As the results of the IEEE S&P 2025 paper [210] drew heavily on the interview study that was conducted by Collins, and my contributions to the results focused more on the app analysis portion, in the results I focus on the research question most pertinent to the modded app analysis (**RQ4**). I include the interview study methods and summaries of its results for completeness, as participants' expectations highlighted in the interview study informed the app analysis.

6.3 Background on WhatsApp and modified Android apps

Here, we provide a background on WhatsApp, modded Android apps, and WhatsApp mods and their features.

6.3.1 WhatsApp

WhatsApp is the world's most popular social messaging app [262] with over two billion users [286]. Messages and calls over WhatsApp have been end-to-end encrypted (E2EE) since 2016 [154], and recently, WhatsApp has introduced additional features related to security and privacy, e.g.,

- Chat lock: Allows users to hide specific chats behind a biometric or password (May/Nov 2023) [288, 289].

- Multiple accounts on one device (Oct 2023) [290].

As we will discuss more in Section 6.4, some participants in our user study described these features (and others) as reasons they used WhatsApp mods over the official WhatsApp (despite the features already being in the official WhatsApp).

Although we do not have insight into the origins of various features in the official WhatsApp, we note that features that have recently been implemented in the official app (e.g., chat lock and allowing multiple accounts on one device) were already present in some WhatsApp mods years before; one website advertised a mod with these features in 2021 [24]. WhatsApp may be implementing some mod features to discourage users from using WhatsApp mods; in its blog post for its new feature allowing multiple accounts to be used (a popular mod feature mentioned by participants in our study), the post ended with a statement encouraging users to *not* use mods to access this feature [290].

6.3.2 Modified Android apps

Android apps are sometimes modified (or "repackaged") to incorporate preferred features into the apps (e.g., a new design for a social media app), to access premium features for free (assuming they are enforced client-side), or to impersonate an app and monetize its users [233, 303]. Several existing tools (including some built into Android Studio) allow decompiling APKs (i.e., Android Packages) to human-readable code to facilitate debugging, and a knowledgeable developer could decompile, modify, and repackage an app in minutes [100]. Using these tools, people decompile the apps, modify the code, and repackage them and sign them using their own digital certificates (i.e., the developer's public-private key pair and other metadata). While there are some benign reasons for modding an app, research has shown that modded apps are a vehicle for malware [71], with one study finding that about 86% of Android malware is from modded apps [304]. More recently Saavedra et al. studied app marketplaces *specifically for modded apps* and through their analysis of over 146 thousand apps, they found that 8-9% of mods were labeled as malicious by VirusTotal [242], mirroring our findings in Section 6.5.2. While we focus on Android apps, and particularly modified WhatsApps, given Apple's recent announcement regarding alternative app marketplaces in response to the EU's Digital Markets Act [34], we expect that modded apps may soon become widespread on iOS devices.

In their literature review on detecting repackaged Android apps, Li et al. found that determining repackaged app provenance is challenging [179]. Given a pair of repackaged apps, there is no straightforward way to determine which app was the "original" app. This means that given a corpus of WhatsApp mods (as we have collected in this study), it is challenging to determine which modded app is the "official" GB WhatsApp and which apps are mods of mods. In a study of six third-party marketplaces, Zhou et al. developed a system (DroidMOSS) to detect repackaged apps at scale [303]. For the approximately 23,000 apps they collected from third-party marketplaces in the US, China, and East Europe, they randomly sampled 200 apps and compared them to another corpus of apps they had from the

official Android Market (i.e., Google Play Store). They found that 5-13% of the apps were repackaged, and the rest were either redistributed from the Android Market or were only available in third-party marketplaces. In our study, we find that popular WhatsApp mods in Kenya are hosted on individual websites rather than third-party marketplaces, potentially posing a lot of security and privacy risks to users.

6.3.3 WhatsApp mods

WhatsApp mods are unofficial, modified versions of WhatsApp developed by third parties. These apps offer additional features beyond what is provided by the official versions of WhatsApp, e.g., the ability to disable incoming calls, and are interoperable with the official versions. We believe that these mods still support E2EE (due to their continued interoperability and some ad-hoc testing of WhatsApp's signature verification feature), but that does not mean that information is not extracted once messages are decrypted client-side. Although the early origins of these mods are not well-documented, they are believed to be descended from an open-source reverse engineering effort called "libwhatsapp" [106] that aimed to create a usable gateway to WhatsApp; this library has been archived and is currently read-only [96]. The original GB WhatsApp page (believed to have originated in Syria during its conflict) was taken down in response to legal threats from Meta (then Facebook) in August 2018 [96]. WhatsApp mods have become very popular in some places in Africa, e.g., Kenya [96]. At the same time, these apps pose potential security and privacy risks as they are mostly distributed through third-party websites or marketplaces or via sharing of the APKs directly (as confirmed by several participants in our interviews), evading app store security measures [239].

Modded app features The WhatsApp mods in our study claimed to provide several features not offered by the official WhatsApp. In this chapter, we explored a subset of the features that have clear security and privacy implications for *interpersonal communication*:

- **Anti-delete:** Often enabled by default, this feature preserves messages and statuses

(ephemeral text or media updates that expire after 24 hours) posted by other users, even if they have been deleted globally.

- **Asymmetric hiding of read receipts:** This feature allows a WhatsApp mod user to receive messages and view messages without the sender being notified that the messages have been read while *the mod user is still notified about the sender reading their messages*. This differs from official WhatsApp where hiding read receipts is symmetric; in the official WhatsApp, either both users can see or both cannot see read receipts [292].

- **Hide status views:** Similar to the previous feature, this allows a WhatsApp mod user to view statuses without their name showing up in the list of status viewers.

- **Freeze "last seen":** WhatsApp mod users can freeze their "last seen" time (e.g., `11:57 am`) to conceal when they are actively using the app.

- **Disable incoming calls:** Mod users can reject incoming calls and *how* the call is rejected, e.g., they can make it appear as if they did not answer after several rings or that they do not have internet connection. In official WhatsApp, one can disable calls from unknown numbers and block calls from specific numbers [293] but not block calls from all numbers.

Several other mod features are available in the official WhatsApp. However, there exists some confusion among participants about which features are only available through the mods. We discuss these misconceptions in Section 6.4.2.

6.3.4 Threat model

The potential security and privacy (S&P) risks from using WhatsApp mods motivate our investigation. In this work, we consider two primary threats to users (both WhatsApp mod and *non*-mod users):

Threats from mod developers There is a security risk to WhatsApp mod users (and the people they communicate with) from malicious or careless developers. A malicious developer could use the mod to compromise users' devices–e.g., to install adware for financial purposes. A careless developer could also insecurely transmit or store users' data.

Threats from modded apps' users People trust that data sharing in WhatsApp will be symmetrical: e.g., if I enable read receipts, someone that I send a message to can only see that I have read a message if they also have read receipts enabled. WhatsApp mods' claimed features break this trust.

6.4 Interview study methods and findings

To understand why and how people use (**RQ1**), propagate (**RQ2**), and perceive (**RQ3**) modified WhatsApps, we first conducted in-depth, semi-structured interviews ($n = 20$) with participants that use modified WhatsApps in Kenya. As mentioned in Section 6.2, the interview study portion of this work was done by my co-author Collins Munyendo (with some assistance from Faith Strong on qualitative analysis).

6.4.1 Methods

Recruitment and demographics To recruit participants, we first employed the "anti-delete" feature (see Section 6.3.3), one of the most common features of WhatsApp mods which is often enabled by default. One of the researchers, who is from Kenya, posted and immediately deleted an advertisement for the study on his WhatsApp status. Five users of WhatsApp mods were able to view and respond to this deleted update, and were recruited to the study. We also posted about the study in some local WhatsApp groups in Kenya. Afterward, we used snowballing [122] to recruit further participants, whereby participants recommended others to participate whom they know use WhatsApp mods. For all participants, we used the "anti-delete" feature to verify they were indeed using WhatsApp mods by sending them a code and immediately deleting it. Only participants that were still able

to view and reply to this code were included.

To be eligible for the study, participants had to be 18 years of age or older, and current users of any WhatsApp mod. While we had an equal number of male and female participants, most participants were young (aged mostly between 25 and 34 years), with varying education levels. Table D.1 has the full demographics of participants.

Interview procedure We first asked participants to indicate all WhatsApps they currently use, as it is possible to use multiple WhatsApps and mods at the same time. We asked these participants to discuss only the most frequently-used mod. We then asked whether they were also aware of the official WhatsApp and, if so, the similarities and differences between their WhatsApp mod and the official WhatsApp. Subsequently, we inquired approximately when participants started using their WhatsApp mod and their primary motivation for doing so. We also asked participants how they learned about and installed their WhatsApp mod, as well as any factors they considered when doing so.

After asking participants about one thing they like and dislike the most about their WhatsApp mod, we asked whether they had recommended the app to any other people. Our next questions centered around trust in the modded and official WhatsApps, including factors that make participants to trust or not trust these apps as well as their developers or owners. We additionally inquired about any concerns that participants have in their WhatsApp mods.

We further asked about participants' update practices for their WhatsApp mods, including why and how they make these updates. We also asked about any challenges they had faced because of using mods. Lastly, we asked whether participants post on their WhatsApp status as well as any privacy controls they have around these updates. The full interview protocol is available in Appendix D.

Data collection We first conducted two mock interviews with two qualitative researchers from our lab via Zoom. Based on their feedback, we added more probes and follow-ups to

some questions, following best practices for qualitative work [76]. Afterward, we conducted three pilots with users of WhatsApp mods in Kenya. These interviews were conducted remotely through WhatsApp audio calls.

We conducted all interviews remotely via WhatsApp audio calls in November 2023. As our interview procedure was only slightly altered following the pilot interviews we included the three pilots in our final analysis, and so in total, we interviewed 20 participants, with interviews lasting 38 minutes on average. We took notes during the interviews, and stopped observing new themes after about 15 interviews; therefore, 20 interviews was likely sufficient for us to reach saturation. All interviews were audio-recorded and conducted in English, one of the official languages in Kenya [218]. Similar to recent studies in Kenya [208, 209], participants were compensated with either 125 minutes of call time or 2 GB worth of internet data. This was directly transferred to participants' phones after the interviews.

Data analysis Using MAXQDA [198], we qualitatively analyzed the interview transcripts [76]. Two researchers began by collaboratively coding one transcript to develop an initial primary codebook, and then used this codebook to independently code three transcripts before meeting to resolve differences and update the codebook. The researchers repeated this process until all transcripts had been coded, regularly meeting to resolve differences, update the codebook, and discuss emerging themes. At the end of this process, the researchers revisited all transcripts to confirm everything had been accurately coded and all updates to the codebook were consistent across all the transcripts. Since the two researchers resolved differences across all the transcripts, there was no need to compute inter-rater reliability [200].

Limitations This study has several limitations. Foremost, our investigations are only limited to Kenya. While Kenya is one of the countries with the highest adoption and usage of WhatsApp mods [162], future work is needed to explore how these apps are perceived and used in other countries and contexts. As is typical for qualitative studies, our sample size was relatively small and young. While we do not claim that our results generalize within

Kenya, we took comprehensive notes during each interview and only ended data collection after we stopped noting new themes. Additionally, these apps are primarily used by young people, as mentioned by several participants in the study, and so we would expect a younger demographic in a sample of such users.

As is typical for paid studies, some participants tried to participate in the study for financial reasons despite not meeting the eligibility. We mitigated this by using the "anti-delete" feature of WhatsApp mods to ensure participants were actually using these apps before conducting interviews. One participant that had been recommended to participate was excluded this way. This study may also suffer from social desirability bias where interviewees try to look more favorable to the interviewer. We tried to mitigate this by telling participants that we were not testing them but only interested in their honest thoughts and practices. Most participants appeared honest, e.g., one admitting that they had used their WhatsApp mod to stalk their partner.

Ethical considerations This study was reviewed and approved by two academic institutions' Institutional Review Boards (IRBs). Participation was voluntary and participants could withdraw from the study at any time without any consequences. We also encouraged participants not to respond to any questions they were not comfortable answering. To minimize any potential risks of unauthorized access of the study data, we did not collect any personally identifying information (PII) from participants during the interviews. Any PII captured or inadvertently disclosed by participants during the interviews was removed during transcription. In Section 6.6.2, we discuss our plans to disclose our findings to WhatsApp mod users and developers.

*6.4.2 Summary of qualitative results (**RQ1**, **RQ2**, & **RQ3**)*

To address **RQ1**, **RQ2**, and **RQ3** , I will now present a summary of the qualitative results from our interview study ($n = 20$) with WhatsApp mod users in Kenya. As counts and percentages could imply generalizability, we primarily use quantifiers such as most and few

Table 6.1: WhatsApps used by participants. Participants could indicate multiple, and the three participants using the official WhatsApp were predominantly using WhatsApp mods.

WhatsApp Name	**App Type**	**No. of Part.**
GB WhatsApp	Mod	18
Official WhatsApp	Official	3
WhatsApp for Business	Official	1
Yo WhatsApp	Mod	1
TM WhatsApp	Mod	1
GB WhatsApp Pro	Mod	1
FM WhatsApp	Mod	1

when presenting the results. For certain results where we do provide counts, we caution against drawing any generalized findings. Our fact-checking of participant sentiments in this section is derived from our feature analysis, described later in Section 6.5.2.

*Common mods and reasons for their use (**RQ1**)*

In response to **RQ1**: *Why and how do people use WhatsApp mods?*, we detail the WhatsApp mods most commonly-used by participants as well as their reasons for using these apps.

RQ1 summary The most common mod participants used was GB WhatsApp (see Table 6.1 for more details). Based on our interview study, several "advanced" features motivate people to use WhatsApp mods; we found that several of them are also supported by the official WhatsApp. This could be due to misinformation on the modded apps' websites (i.e., checklists showing the features that the mods have that the official app does not [294]) or just

a lack of knowledge, particularly about features recently added to the official WhatsApp.

*Propagation of WhatsApp mods (**RQ2**)*

In response to **RQ2**: *How do users describe learning about, acquiring, and updating WhatsApp mods?*, we describe how participants acquire, update, and distribute mods, as well as the considerations and trade-offs they make when using them.

RQ2 summary Overall, we find that recommendations from friends, family, and work colleagues are the most common way in which participants learn about, install, and distribute WhatsApp mods. As most of these mods are not available on the Play Store, participants mostly install these apps from APKs directly shared with them, making them susceptible to installing potentially malicious apps. When installing these apps, most participants do not consider any specific factors as they are more motivated to get the app. Lastly, a majority of participants regularly update their WhatsApp mods to prevent the apps from expiring; the frequent updates that mods require (to avoid their users being blocked) serve to normalize risky security behaviors.

*Challenges, practices, and mental models (**RQ3**)*

Here, we address responses to **RQ3** regarding WhatsApp mod users' concerns, challenges, and practices with mods as it relates to their security and privacy mental models.

RQ3 summary Overall, we find that many participants struggle with obtaining as well as updating their WhatsApp mods due to the absence of these apps on the Play Store. Nine participants indicated they trust the official app more while eight participants trusted the mod more, with three having the same level of trust for both applications. Common reasons for trusting the official app included the app being original, its availability on the Play Store, its simplicity, and no past incidents. Reasons for the distrust of the official WhatsApp included the lack of privacy controls and privacy more generally as well as lack of advanced

features compared to the mods. We also find some contradiction in participants' practices and expectations when using the mods, with participants often motivated by the mod features such as *anti-delete* but at times displeased when others equally use these features "against" them. Further, several participants detail experiences where the participants themselves or others they know have used WhatsApp mods to either spy on, or stalk others. Interestingly, almost all participants believe that WhatsApp mods offer them more control over their data compared to the official app.

6.5 Contextualizing users' perceptions of mods

Throughout the user study, participants made several claims about the mods' behavior. While some of these claims can be evaluated with a quick online search (e.g., does the official WhatsApp offer a specific feature), broader claims about the mods' unique features, permissions requested, and trustworthiness demand a broader, more systemic evaluation. Specifically, participants believed that (1) WhatsApp mods offer unique S&P features that hide information from the person with whom they are communicating, (2) WhatsApp mods request similar permissions as the official WhatsApp, and (3) WhatsApp mods are more trustworthy than the official app; almost half of participants expressed this final sentiment.

To investigate how participants' expectations of WhatsApp mods either align or misalign with the apps' actual behavior (**RQ4**), we analyzed common WhatsApp mods, particularly those most frequently used by participants in the user study (see Section 6.4). For a subset of the WhatsApp mods, we conducted a *feature* analysis and manually validated that they offered the features they claimed. Additionally, we analyzed the output of VirusTotal [25]—an online software security analysis platform—to understand the permissions these apps requested and if they contained malware.

6.5.1 Methods

Identifying and selecting WhatsApp mods. As WhatsApp mods are generally not available on the Play Store, it is challenging to find the "official" source for these apps. We

chose to focus on "GB WhatsApp" because this name is associated with the most popular mod [162], and almost all participants in our interview study (described in Section 6.4) used it. A cursory search for "GB WhatsApp" will yield numerous websites that all claim to have GB WhatsApp for download on their pages. We saved the top ten search results and downloaded their APKs to our local machine in December 2023 (shortly after we conducted the interviews).

For the purposes of our study, we wanted to understand the *popularity* of WhatsApp mods we downloaded. Given the difficulty of both quantifying and trusting the number of downloads outside of official app marketplaces, we decided to use the Chrome User Experience Report (CrUX) [114]—an internet measurement report that among other metrics outputs a *popularity* metric for domains that surpass an undisclosed minimum threshold of unique visits. We used CrUX for internet traffic in Kenya during the month of December 2023 to determine the popularity of websites that we found hosting WhatsApp mods. Using this approach, we found that multiple websites hosting WhatsApp mods (e.g., https://gbapps.org.pk) were in the top 1000 of websites visited in Kenya during the time of our study. In total, we collected 14 APKs (ten from the WhatsApp mods we downloaded, three from these apps' updates, and the official WhatsApp from the Google Play Store), shown in Table 6.2.

Feature analysis To test whether the features described in Section 6.3.3 and by participants functioned as described, we developed a *testing procedure* that we walked through for five mods in our corpus and the official WhatsApp (see Appendix D.2 for the full procedure). This procedure involved a series of steps on a *test phone* (a Google Pixel 5a with the target app installed) and a non-test phone (with the official WhatsApp installed). For example, when the "anti-delete" feature was enabled, we deleted a message on the non-test phone and observed the test phone to ensure the message was still visible. We initially chose to use the top five mods in our search results (MODs 1-5). We attempted to use MOD5, but the app was blocked by Google Play Protect when we attempted to sideload it using ADB (for reasons that we discuss in Section 6.5.2). To simulate actions that a typical user would

Table 6.2: Metadata of the apps we analyzed. The thumbprint is the SHA1 hash of the digital certificate used to sign the APK. The CrUX rank indicates that a website was in the Top X websites during December 2023 in Kenya.

App ID	Package Name	Thumbprint	SDK	APK SHA256	CrUX Rank	Source
MOD1	com.gbwhatsapp	61ed377e...	33	50769499...	1000	https://gbapps.org.pk
MOD2	com.universe.messenger	c8df88cd...	33	ea37bf76...	5000	https://www.gbwhatsapp.chat
MOD3	com.universe.messenger	c8df88cd...	33	e5827f17...	5000	https://www.gbwhatsapp.download
MOD4a	online.whatsticker	bea2d1d9...	33	e71a72cb...	5000	https://allwapk.com
MOD4b	online.whatsticker	bea2d1d9...	33	f1203e04...	5000	https://allwapk.com
MOD4c	com.aerowtsapp	bea2d1d9...	33	278c1435...	5000	https://allwapk.com
MOD5	com.gbwhatsapp.sofid	e07080ed...	33	512958b7...	1000	https://gbapps.net
MOD6a	com.universe.messenger	c8df88cd...	33	ea37bf76...	50000	https://www.whatspro.org/
MOD6b	com.universe.messenger	c8df88cd...	33	d6551b78...	50000	https://www.whatspro.org/
MOD7	com.gbwhatsapp	61ed377e...	33	50769499...	1000	https://androidwaves.com
MOD8	com.gbwhatsapp	e509c3c1...	33	fbed8a41...	1000	https://gbwasap.com/
MOD9	online.whatsticker	bea2d1d9...	33	f1203e04...	50000	https://gbwhatsapp.en.malavida.com
MOD10	com.gbwhatsapp	61ed377e...	29	352ae77c...	50000	https://gbwatsapp.download/
Official	com.whatsapp	38a0f7d5...	33	2976510d...	N/A	Google Play Store

likely take (i.e., *not* disabling Google Play Protect), we proceeded to use MOD6a instead of MOD5 for our analysis. We went through the procedure manually twice for each of the five apps. During this procedure, we also collected network traffic from the test phones to examine the advertising and tracking domains contacted by the WhatsApp mods and their usage of TLS compared the official WhatsApp.

VirusTotal reports To understand (1) the permissions that WhatsApp mods requested relative to the official WhatsApp and (2) whether WhatsApp mods exhibit malicious or undesired behavior, we uploaded the apps to VirusTotal. VirusTotal is a comprehensive, multi-tool scanning device widely used in industry and research [147, 264, 306] and maintained by Google Cloud's Chronicle Security Operations [81]. VirusTotal aggregates several anti-virus (AV) engines. Prior work [43] has shown that AV engines may be unreliable; we follow the approach used by Wang et al. [281] and consider apps as *malicious* if at least 10 AV engines (a number found to be a robust threshold [43, 148, 301]) classify them as such. VirusTotal also outputs information about APKs using Androguard (e.g., permissions).

*6.5.2 Results (**RQ4**)*

The WhatsApp mod ecosystem changes rapidly and is at times unreliable, with several apps 1) not working or 2) requiring an update within a short period of time For example, nine days after we initially downloaded MOD4a, the app would not function and required an update. After following the links provided by the app, we downloaded MOD4b and ran it through our procedure (Section 6.5.1). When running a second procedure with MOD4b 11 days later, we were prompted to install another update; since this update was not mandatory like the previous one, we simply ignored it and installed the app (MOD4c) after we finished the procedure. This experience mirrors complaints we observed from participants in the interview study about the difficulty and frequency of updates; we were prompted to manually update the app twice in 20 days.

For the S&P features we tested, we found that all the WhatsApps mods supported them The S&P features claimed by WhatsApp mods that we focused on were: anti-delete, hide message read receipts, hide status views, freeze "last seen," and disable incoming calls.[1] As described in Section 6.5.1, we designed a procedure to verify that these features functioned as they were advertised, and for the features we tested, we found that all the WhatsApp mods indeed supported them. Disabling incoming calls caused the modded WhatsApp to receive a notification of a missed phone call, but the phone did not ring. Hiding read receipts and hiding status views also functioned properly and as expected, but the features varied in the amount of information they revealed to the WhatsApp mod user.

For example, when we tested the "anti-delete" feature, we observed that after a message was deleted from the official WhatsApp, the message was still visible in the WhatsApp mod, and the message had a symbol (⊘) added to its display indicating that it had been deleted. However, when we observed a status on the test phone after it had been deleted on the non-test phone, we did not observe any visual indicators. This may explain how people using

[1]As we discuss in Section 6.3.3, the ability to hide read receipts and hide status views in mods is *non-reciprocal*, unlike in the official WhatsApp.

mods accidentally leak that they are using a mod—by responding to a deleted status that they are not aware was deleted.

Freezing "last seen" time also functioned properly but with a surprising side effect: the mod user is unable to view other users' "last seen" time. This may be due to the way the feature is implemented, and it introduces a trade-off for the WhatsApp mod user—by hiding information (your activity) from other users, you also lose information about others.

Although the majority of the permissions are similar, WhatsApp mods requested permissions that allow them to take privileged actions, such as editing system settings or drawing on top of other apps Some participants described not being concerned about permissions requested by the WhatsApp mods because they were the same as or similar to the official WhatsApp. However, there were seven permissions requested by WhatsApp mods that were not requested by the official WhatsApp, including one dangerous permission (`ACTIVITY_RECOGNITION`), one permission that is ignored for third-party apps (`MOUNT_UNMOUNT_FILESYSTEMS`), two normal permissions (`QUERY_ALL_PACKAGES`, `EXPAND_STATUS_BAR`), and three *signature* permissions that have a distinct approval flow (in Android API level 23+) and navigate the user to a separate screen (`WRITE_SETTINGS`, `SYSTEM_ALERT_WINDOW`, `MANAGE_EXTERNAL_STORAGE`). Respectively, these three permissions could enable a mod to change device settings to conceal its behavior, to allow the app to appear on top of other apps and change the way other apps appear, or to read, change, or delete files in storage. We present a list of extra permissions for each mod in Table 6.3.

Two WhatsApp mods were malicious (AV-count $>$ 10), and were classified as trojans As presented in Table 6.4, two of the apps (MOD5 and MOD8) contained the Triada Trojan [256], a trojan that collects data from devices, downloads malicious payloads, and is known to be distributed via WhatsApp mods [52, 201]. Several other WhatsApp mods were considered adware by some AV engines, but there was not enough consensus to consider them malicious.

Table 6.3: Mod permissions that are not requested by the official WhatsApp.

Permissions	ACTIVITY_RECOGNITION	EXPAND_STATUS_BAR	MANAGE_EXTERNAL_STORAGE	MOUNT_UNMOUNT_FILESYSTEMS	QUERY_ALL_PACKAGES	SYSTEM_ALERT_WINDOW	WRITE_SETTINGS
MOD1						X	
MOD2	X	X	X	X	X	X	X
MOD3	X	X	X	X	X	X	X
MOD4a						X	
MOD4b						X	
MOD4c			X			X	
MOD5						X	
MOD6a	X	X	X	X	X	X	X
MOD6b	X	X	X	X	X	X	X
MOD7						X	
MOD8						X	
MOD9						X	
MOD10			X			X	

Domain analysis Modded apps contact several domains that the official app does not, including tracking and advertising domains. For all the apps we analyzed, there were 52 total domains visited, with the apps visiting between zero (only the official WhatsApp) and 40 advertising/tracking domains as determined by uBlock Origin's ad/tracker list [183]. Within

these 40 domains, the most common domains contacted by the mods were `flurry.com` (MODs 2,3,4b,6a), `vungle.com` (MODs 2,3,4b,6a), and `doubleclick.net` (MODs 2,3,6a). While carrying out the procedure, we observed ads in 4/5 of the modded apps that we manually tested (MOD2, MOD3, MOD4b, and MOD6a), as might be expected from the advertising/tracking domains observed in their network traffic. On some occasions, these ads had fake 'X' buttons that would redirect the user to the Google Play Store rather than closing the ad. Table D.2 in Appendix D contains the domains visited by each app in our procedure.

TLS utilization Similarly to official WhatsApp, almost all (98-100%) of the captured network traffic in mods is encrypted via TLS. The few requests using HTTP were GET requests to app-specific domains (e.g., alexmods.com) for HTML pages with recent app updates.

RQ4 summary Our analysis of the mods mirrors participants' frustration with the ephemerality of the mods and confirms that features such as "anti-delete" mentioned by most participants function as described. However, common WhatsApp mods used by participants pose various risks, with all mods requesting more permissions than the official app and two mods classified as malicious by VirusTotal. These findings are troubling, given that almost half of participants trust WhatsApp mods more than the official WhatsApp.

6.6 Discussion and conclusion

In this section, we reflect on our findings and contextualize the risks posed by WhatsApp mods. We also provide some recommendations for relevant stakeholders.

6.6.1 Mod Implications and Takeaways

Mod users' security and privacy (S&P) mental models center around other people, and not developers. Throughout our study in Kenya, we observed that WhatsApp

Table 6.4: Malware classificiation. AV-count is the no. of anti-virus engines classifying an app as malicious.

App ID	AV-count	VirusTotal Label
MOD1	6	andr/wamod
MOD2	5	adware.hiddenad
MOD3	4	adware.hiddenad
MOD4a	0	unknown
MOD4b	2	unknown
MOD4c	3	unknown
MOD5	14	trojan.triada/bankbot
MOD6a	5	adware.hiddenad
MOD6b	3	unknown
MOD7	6	andr/wamod
MOD8	12	trojan.triada/frtr
MOD9	2	unknown
MOD10	6	pua
Official WhatsApp	0	unknown

mod users' S&P mental models revolved around their social circles i.e., their friends and family, and not the developers of the mods. Similar to Facebook users in South Africa [236], most WhatsApp mod users were more worried about what they could see from their friends, including deleted content as well as what their friends could see about them; and not the mods and their potential security and privacy issues. In fact, most participants believed the mods are more secure than the official WhatsApp because of the level of control around their privacy that the mods afford them. Even those that had concerns about the mods were willing to overlook these concerns to have the privacy controls from the mods. At the same time, we found that these apps pose various S&P risks, with some of them containing

malware.

Due to the risks posed by the mods, we suggest that official WhatsApp should prioritize feature updates that give users more control e.g., by allowing them to disable calls, to dissuade users from using the mods. In fact, several participants indicated they would prefer to use the official app if it afforded them more privacy controls. WhatsApp could pay close attention to the mods and implement some of their favourable privacy controls (and may already do so, albeit slowly), and/or elicit feedback from users. However, we caution against incorporating adversarial features such as "anti-delete" as they ultimately harm users' privacy, e.g., when they accidentally share or post sensitive information.

Security advocacy is a promising way of teaching good security and privacy practices While WhatsApp mods are often not on official app marketplaces, they have still managed to grow in popularity. For instance, GB WhatsApp is the second most used social messaging application in Africa, behind only the official WhatsApp [162] and ahead of Facebook Messenger. In our study, we found that most mod users learn about and distribute these apps through their social circles of friends, family members, and work colleagues, similar to WhatsApp mod users in Pakistan [212]. There is perhaps an opportunity to leverage these social connections and contacts to spread good security and privacy practices through security advocacy. As prior work in Kenya [209] notes that people often get S&P advice from "local experts" such as managers at public computing facilities such as cybercafes, targeting and teaching these "experts" good security and privacy practices could have widespread impact. Exploring the efficacy and feasibility of such initiatives is a promising direction of future research.

Some participants have ambivalent attitudes when using WhatsApp mods Ironically, we noticed some contradictions between some participants' preferences for their own S&P in comparison to others. While most participants were drawn to mods because of features such as "anti-delete," several expressed frustration and anger when others were able to

view things they had deleted. Some even admitted that the mods have ruined privacy. Similar to the "privacy paradox", it might be interesting to explore the prevalence and root causes of such seemingly contradictory attitudes, especially for chat applications such as WhatsApp where settings are often symmetrical, i.e., you can see other peoples' read receipts *only* if you enable your own read receipts.

The use of WhatsApp mods poses various security and privacy (S&P) risks, both to mod users and non-users As our results show, people who use mods are potentially exposing themselves to malware by installing them. Two of the WhatsApp mods were clearly malicious with several other mods displaying dubious behaviors. Even if the apps are not malicious at install-time, because of the frequent and dynamic nature of their updates, users cannot be sure that a later version of a WhatsApp mod will not include a malicious payload. Beyond the more serious malware risk, there is a broader privacy risk from the advertising/tracking the apps do; these could ostensibly monetize users by sharing data about them with advertisers and analytics companies. Therefore, we caution users do their due diligence before choosing to use mods. One simple way to do this could be assessing the mod's APK via VirusTotal before installing it.

Non-users of WhatsApp mods, on the other hand, are placed in a difficult situation. By using the safer, official WhatsApp downloaded from the Google Play Store, they are at an informational disadvantage and on the lower end of a power imbalance. If they are aware of the existence of modded apps and their features, they know that they cannot trust read receipts or "last seen" times, and they know that there is a risk that anything they send and later delete could be retained by a modded app user. If they are not aware of others' usage of modded apps, there is more severe informational asymmetry. WhatsApp mod users may deceive or manipulate them using the modded app features, and the non-users will trust everything WhatsApp's UI tells them. In that way, non-users who are not aware of mods may be more susceptible to deception and scams. One potential remedy to this could be WhatsApp informing non-mod users whenever they are communicating with a

mod user [212], alongside the potential implications of that.

6.6.2 Responsible Disclosure

Websites hosting WhatsApp mods. By reviewing common WhatsApp mods via Virus Total, we found two of them to be malicious and containing malware. There is a chance that the websites hosting these WhatsApp mods took the APKs from other sites not in our study and repackaged them (unaware that they contained malicious components). However, there is also a chance that these websites knowingly host malicious apps. At the very least, we assume the people behind these websites know that they are impersonating the official WhatsApp and in some cases monetizing users. Nevertheless, we reached out to the two websites hosting modded APKs classified as malicious and notified them of our findings. We recommended that they remove all malicious APKs from their websites and upload any new APKs to VirusTotal before hosting them.

Participants using WhatsApp mods While we conducted data collection for the interviews and the app analysis simultaneously, we did not analyze the mods until after the interviews were completed. This means that we did not get the chance to tell participants that these apps potentially contained malware. Moreover, because we could not know the specific APKs that the participants had on their devices, we would not have been able to immediately confirm if *their* modded app contained malware.[2] In the end, we drafted a message to all participants summarizing our findings, and suggesting that they upload their APKs to VirusTotal if they want to check if their mods are malicious. Working with our IRBs, we have appropriately re-contacted all participants.

[2]Two apps with identical package names cannot be installed on a device. The Google Play Store does not contain apps with identical package names, so they are often used as identifiers for apps. However, since modded apps are sideloaded, their package names can conflict with those of existing apps. Due to these factors, unless participants sent us a hash of their APK, we could not identify which specific mod they were using.

6.6.3 Recommendations and Conclusion

Lessons and recommendations. Our results are directly useful to various stakeholders, including end-users, developers, WhatsApp, and application markets. For WhatsApp mod users and other users generally, we caution against using modded apps without doing due diligence, as some of these apps are potentially malicious and may share users' personal information with advertisers and other third parties. For developers hosting these apps, especially mods of other mods, we also emphasize the need for due diligence; copying or downloading other mods directly without closely inspecting their codebase could lead to developers unintentionally hosting and sharing potentially harmful applications. For WhatsApp, we believe there is an opportunity to stay up to date and incorporate some of the favourable mod features into the official WhatsApp. After all, several participants indicated they would prefer to use the official app if it afforded them more privacy controls. For third-party marketplaces, it might be useful to regularly scan their stores with the goal of detecting and removing apps that are potentially malicious.

To use or not to use ... that is the question We decline to answer the question: *should people use WhatsApp mods?* Instead, we attempt to elucidate the good, the bad, and the ugly of mods, from the users' perspective. Mod users gain features that they value that the official WhatsApp does not offer. For example, the ability to download statuses allows for media portability, and the ability to block all incoming calls (while also controlling how the caller perceives the blocking) could be considered an anti-spam or an anti-harassment feature. While other features (e.g., freezing "last seen") facilitate information asymmetry, this is not necessarily a bad thing—particularly if the asymmetry benefits someone with less power (e.g., the example from Naveed et al. [212] in Section 2.3.2). While there are risks that, without security checks in place, this ecosystem can facilitate the spread of malware, WhatsApp mods also have clear utility (and sometimes privacy and safety) benefits for end-users.

6.7 Technology & new power imbalances

This work highlights a power imbalance between users and other users. People who use the official WhatsApp are at a disadvantage, and they may likely be unaware of that disadvantage. Mod users have access to more information and more control of that information than the official WhatsApp users. This informational asymmetry is a form of power to which official WhatsApp users did not consent. This power imbalance creates a negative feedback loop, in which official WhatsApp users are incentivized to use mods in order to reduce this imbalance. Users will have to decide for themselves if the increased power they get from using mods is worth the risk of malware; this is a not a decision that they should have to make.

Chapter 7

CONCLUSION AND REFLECTIONS

In this dissertation I have presented four cases studies analyzing how technology amplifies existing (Chapters 3 to 5) or generates novel (Chapter 6) power imbalances. Table 7.1 presents a summary of these case studies, including the pre-existing power imbalance, the relevant technology, what the impacts we uncovered in our research were, and potential remedies. These case studies fall under three categories of power imbalances: between users and government entities (Chapters 3 and 4), users and corporations (Chapter 5), and users and other users (Chapter 6).

Chapter 3 showed us how electronic monitoring smartphone apps increase surveillance and the compliance burden for people under community supervision. In Chapter 4, we learned how `BI SmartLINK` can cause migrants attempting to immigrate to the U.S. to lose access to housing, jobs, and community, in part because of a lack of transparency about which data the app collects (and when exactly it collects those data) and the threats it poses to people around migrants. Chapter 5 taught us how voice interfaces can be used to manipulate users in novel ways, and Chapter 6 highlights the risks that WhatsApp mods pose to both users and non-users (i.e., official WhatsApp users) of WhatsApp mods.

7.1 Comparisons of power imbalances and their impacts

In Chapters 3 to 6, I describe power imbalances in which users (or a subset of users) are in a disadvantageous position. How disadvantaged they are depends on the specific type of power imbalanced discussed, and the negative consequences of the tech-impacted power imbalances presented in this dissertation vary widely.

	Existing power imbalance	**Relevant technology**	**Impacts of the technology**	**Potential remedies**
Chapter 3: Electronic monitoring apps	Surveillance under community supervision	Smartphone apps	More sensitive data access; more supervision conditions to follow; malfunctions may set up users for failure	App marketplaces or regulators could ban EM apps
Chapter 4: Experiences with immigration surveillance	Surveillance from ICE; challenges of migrating into the U.S.	BI SmartLINK (and ankle monitors)	Loss of housing, jobs, and community; more intense surveillance	Discontinuing usage of the app; A transparency tool to answer questions
Chapter 5: Deceptive design patterns	Corporations control how users interact with technology and the choices they have	Voice interfaces (e.g., smart speakers)	Susceptibility to deception; more potential for negative privacy or financial impacts	Include voice interfaces in regulation of deceptive design
Chapter 6: Modified versions of WhatsApp	-	Modified WhatsApps	Misplaced trust; increased potential for manipulation	WhatsApp implements more of the (safe) mod features.

Table 7.1: A summary of the chapters of this dissertation and their contributions

7.1.1 Users and government entities

When it comes to users and government entities, the technologies discussed in this dissertation increase the power of the state to surveil and control users more intensely. Moreover, the groups of users in Chapters 3 and 4 (people under community supervision and migrants)

are already marginalized and have constitutional rights infringed upon [253, 284, 300]. This means that typical harms arising from surveillance to laypeople—such as chilling effects, discrimination, coercion, and selective enforcement [238]—can carry higher risks to users in these contexts (including incarceration or deportation). Users facing this type of power imbalance are coerced to use these technologies, rather than accepting them with informed consent. In this context, it is unclear if informed consent would even be possible absent coercion (e.g., if someone wanting to manage their alcoholism opted into using an electronic monitoring app with a Bluetooth breathalyzer). The power imbalance means that government entities may always have more information about the technology, its data collection capabilities and practices, and who has access to that data.

7.1.2 Users and corporations

This reflection regarding a lack of informed consent could also be extended to the power imbalance between users and corporations. Corporations control users' choice architecture (e.g., their privacy controls). In the case of voice interfaces, because they may be one of many interfaces a platform uses, users may be unaware of what privacy controls are available to them and how to access them. Moreover, users may be unaware that alternatives to the choice architecture they have been presented (e.g., not collecting or sharing certain data) are possible. Arguably, it is difficult to say that consent in this scenario can be truly informed.

Although users also face *commercial* surveillance [1] from corporations, the harms resulting from this surveillance are largely related to users' finances and privacy. This distinction is increasingly muddled, as stories emerged about government entities buying commercial surveillance data from corporations [271], or corporations just give data away directly to governments [230]. Additionally, if governments want to target particular groups of people and control the judiciary, they can request commercial surveillance data from corporations through court orders [88, 296]. This is all to say: although the consequences of the power

[1] "Commercial surveillance is the business of collecting, analyzing, and profiting from information about people," per the U.S. Federal Trade Commission [8].

imbalance between users and corporations may seem less severe than those between users and government entities in the chapters presented in this dissertation, the reality is more nuanced.

7.1.3 Users and other users

Lastly, the power imbalance between users and other users can result in harms to interpersonal security & privacy. The informational asymmetry resulting from WhatsApp mods' features could be used for deception and manipulation. For example, the official WhatsApp allows users to set media messages as "view once" (with screenshot blocking) and supports disappearing messages [287, 291]. However, when someone is using a WhatsApp mod, the "anti-delete" feature nullifies both features; a "view once" image can be viewed more than once, a screenshot can be taken, and disappearing messages are retained. One can imagine a situation in which a WhatsApp mod user may ask an official app user to send them a nude (or otherwise sensitive image), and the official app user sends a photo with the "view once" setting enabled, believing that the image will be deleted after the other user views it. Unbeknownst to the official app user, the mod user keeps the image and may use it for future exploitation or abuse. While storing nudes of other people received on one's device is not uncommon (Geeng et al. found that 62% of people who had engaged in sexting said that they stored photos of others [124]), in this scenario it is done without consent and with the official app user trusting that the person to whom they sent the photo will be *unable* (as enforced by the app) to screenshot or save the original photo on the device receiving the photo. As we can see, while these interpersonal harms may not rise to the level of incarceration or deportation, they can have severe reputational and psychological harms for users on the lower end of this power imbalance.

7.2 Looking forward

In this dissertation I show how researchers can use techniques from usable security & privacy research (e.g., qualitative methods, threat modeling, mobile app analysis) to how technology

exacerbates or creates power imbalances between users and a variety of entities. These methods enable me to mitigate these power imbalances by increasing transparency, both (1) for users regarding how technology functions and (2) for developers and policymakers about technologies' impacts on users. Improving transparency for users enables them to make informed decisions regarding how they interact with a given technology (even if they are unable to opt-out). For developers and policymakers, understanding people's experiences using technologies can help them better design or better regulate technologies to reduce their harms to users.

In future work I aim to focus on empowering users on the lower end of power imbalances by building tools to answer their questions in an automated fashion. Users often have questions regarding if and how specific applications on their smartphones exfiltrate data from their devices. For example, they may be curious if an app (e.g., `BI SmartLINK`) that requests the "Contacts" permission sends their entire contacts list to its servers. I am developing an automated pipeline for answering similar frequently asked questions for users regarding smartphone app behavior. This will involve mapping specific questions (which we solicit via user studies) to API/function calls and/or dynamic behavior. After these mappings are determined, I will pull the most recent version of the app and present user-facing answers to these questions. I begin with a case study of `BI SmartLINK` and attempt to answer the questions raised by participants in Chapter 4. Through this work, I aim to continue trying to mitigate technologies' negative impacts on power balances and one day, hopefully, conduct research that will empower users to imagine and create better future worlds rather than mitigating existing harms.

BIBLIOGRAPHY

[1] BI SmartLINK® Privacy Policy - BI Incorporated. URL: https://web.archive.org/web/20220328202005/https://bi.com/bi-smartlink-privacy/.

[2] Cal. Civ. Code §§ 1798.100-.199 (West 2018).

[3] Cal. Civ. Code §§ 1798.140(1) (West 2018).

[4] Community Justice Exchange et al v. U.S. Immigration and Customs Enforcement et al. URL: https://www.law.berkeley.edu/case-project/alternative-detention-programs-foia/.

[5] COMMUNITY JUSTICE EXCHANGE, JUST FUTURES LAW, MIJENTE SUPPORT COMMITTEE v. U.S. IMMIGRATION & CUSTOMS ENFORCEMENT and U.S. DEPARTMENT OF HOMELAND SECURITY. URL: https://www.law.berkeley.edu/wp-content/uploads/2022/04/22cv2328_ECF1_Complaint.pdf.

[6] Featured Issue: ICE's Alternatives to Detention Program. URL: https://www.aila.org/library/featured-issue-ice-alternatives-to-detention-program.

[7] Freedom of Assembly and Petition: Overview. URL: https://www.law.cornell.edu/constitution-conan/amendment-1/freedom-of-assembly-and-petition-overview.

[8] FTC Explores Rules Cracking Down on Commercial Surveillance and Lax Data Security Practices. URL: https://www.ftc.gov/news-events/news/press-releases/2022/08/ftc-explores-rules-cracking-down-commercial-surveillance-lax-data-security-practices.

[9] NYCLU, ACLU Challenge Ravi Ragbir Arrest, Cite ICE Targeting of Activists. URL: https://www.nyclu.org/press-release/nyclu-aclu-challenge-ravi-ragbir-arrest-cite-ice-targeting-activists.

[10] Office of Privacy and Civil Liberties — E-Government Act of 2002. URL: https://www.justice.gov/opcl/e-government-act-2002.

[11] Privacy Impact Assessment for the Alternatives to Detention (ATD) Program. URL: https://www.dhs.gov/sites/default/files/2023-08/privacy-pia-ice062-atd-august2023.pdf.

[12] Smartphone Monitoring. URL: https://bi.com/mobile-monitoring/.

[13] Surveillance Self Defense: Threat Modeling. URL: https://www.eff.org/document/surveillance-self-defense-threat-modeling.

[14] United States v. Jacobsen, 466 U.S. 109, 129-30. 1984.

[15] United States v. Knights, 534 U.S. 112, 119. 2001.

[16] Samson v. California, 547 U.S. 843, 848-49. 2006.

[17] Riley v. California, 573 U.S. 373, 401-02. 2014.

[18] United States v. Barnett, 415 F.3d 690, 691-93. 2015.

[19] Carpenter v. United States, 138 S. Ct. 2206, 2221. 2018.

[20] A.N. v. Ind. Dep't of Child Servs., No. 18A-JT-2147, 2019 Ind. App. Unpub. LEXIS 883, at *4, *14. 2019.

[21] Aspects of voice usage. https://www.elcomblus.com/aspects-of-voice-usage/, Dec 2020.

[22] Poll shows strong cross-ideological support for dramatically reducing jail and prison populations to slow the spread of coronavirus. http://thejusticecollaborative.com/2020/03/new-report-poll-jail-prison-coronavirus/, Mar 2020.

[23] Ndiaye v. Adducci, No. 4:19CV712, 2020 U.S. Dist. LEXIS 252421. 2020.

[24] GB WhatsApp Pro v14.0.0 Apk Download for Android (Official) — web.archive.org. https://web.archive.org/web/20211221152909/https://gbwhatspro.com/, 2021.

[25] Virustotal. https://www.virustotal.com/gui/home/upload, 2024.

[26] Jacob Aagaard, Miria Emma Clausen Knudsen, Per Bækgaard, and Kevin Doherty. A Game of Dark Patterns: Designing Healthy, Highly-Engaging Mobile Games. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, CHI EA '22, pages 1–8, New York, NY, USA, April 2022. Association for Computing Machinery. doi:10.1145/3491101.3519837.

[27] Rediet Abebe, Solon Barocas, Jon Kleinberg, Karen Levy, Manish Raghavan, and David G. Robinson. Roles for computing in social change. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 252–260, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3351095.3372871.

[28] ACLU. 91 percent of americans support criminal justice reform, aclu polling finds. https://www.aclu.org/press-releases/91-percent-americans-support-criminal-justice-reform-aclu-polling-finds, Nov 2017.

[29] California Consumer Privacy Act. California consumer privacy act (final text of proposed regulations), 2020. URL: https://www.oag.ca.gov/sites/all/files/agweb/pdfs/privacy/oal-sub-final-text-of-regs.pdf.

[30] California Privacy Rights Act. California privacy rights act, 2020. URL: https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.

[31] Philip E Agre. Real-time politics: The internet and the political process. *The information society*, 18(5):311–331, 2002.

[32] Jasmine Aguilera. Privacy Advocates Sue Over Immigrant Tracking Program. URL: https://time.com/6167467/immigrant-tracking-ice-technology-data/.

[33] Tousif Ahmed, Roberto Hoyle, Patrick Shaffer, Kay Connelly, David Crandall, and Apu Kapadia. Understanding physical safety, security, and privacy concerns of people with visual impairments. *IEEE Internet Computing*, 21(3):56–63, 2017. doi:10.1109/MIC.2017.77.

[34] Peter Ajemian. Apple announces changes to iOS, Safari, and the App Store in the European Union — apple.com. https://www.apple.com/newsroom/2024/01/apple-announces-changes-to-ios-safari-and-the-app-store-in-the-european-union/, January 2024.

[35] Hamed Aleaziz and Paulina Villegas. Trump Shuts Down CBP One App, Signaling the Start of His Immigration Crackdown. URL: https://www.nytimes.com/2025/01/20/us/politics/trump-shuts-down-migrant-entry-app-cbp-one.html.

[36] Amazon Alexa. Customize alexa voice design to create your own voice experience, 2022. URL: https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper.html.

[37] Sara Alsherif. Government does the bare minimum to update the Immigration Exemption. URL: https://www.openrightsgroup.org/blog/government-does-the-bare-minimum-on-the-immigration-exemption/.

[38] Aly Panjwani and Hannah Lucal. Tracked & Trapped: Experiences from ICE Digital Prisons. URL: https://notechforice.com/wp-content/uploads/2022/05/TrackedTrapped_final.pdf.

[39] Aly Panjwani and Julie Mao. ICE Digital Prisons: The Expansion of Mass Surveillance as ICE's Alternative to Detention. URL: https://web.archive.org/web/20240415162858/https://www.flipsnack.com/justfutures/ice-digital-prisons-1u8w3fnd1j/full-view.html.

[40] Andrew Hursh. UN rights expert defines 'psychological torture' in new report. URL: https://www.jurist.org/news/2020/02/un-rights-expert-defines-psychological-torture-in-new-report/.

[41] Antje. Power and Power Mapping: Start Here. URL: https://commonslibrary.org/power-and-power-mapping-start-here/.

[42] Chaz Arnett. From Decarceration to E-carceration. *Cardozo L. Rev.*, 41:641, 2019.

[43] Daniel Arp, Michael Spreitzenbarth, Malte Hubner, Hugo Gascon, Konrad Rieck, and CERT Siemens. Drebin: Effective and explainable detection of android malware in your pocket. In *Ndss*, volume 14, pages 23–26, 2014.

[44] Peter Bacchetti. Current sample size conventions: Flaws, harms, and alternatives. 8(1):1–7. URL: https://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-8-17, doi:10.1186/1741-7015-8-17.

[45] Patricio G. Balona. More eyes being fixed on volusia-flagler juveniles on probation. https://www.news-journalonline.com/news/20170625/more-eyes-being-fixed-on-volusia-flagler-juveniles-on-probation, Jun 2017.

[46] Chelsea Barabas. Refusal in Data Ethics: Re-Imagining the Code Beneath the Code of Computation in the Carceral State. 8(2):35–57. URL: https://estsjournal.org/index.php/ests/article/view/1233, doi:10.17351/ests2022.1233.

[47] Joseph Bartolotta. Usability testing for oppression. 7(3):16–29. URL: https://dl.acm.org/doi/10.1145/3321388.3321390, doi:10.1145/3321388.3321390.

[48] Nikki Trautman Baszynski. States should abolish technical violations of probation and parole. *The Point by the Appeal*, Apr 2021.

[49] Rosanna Bellini, Emily Tseng, Noel Warford, Alaa Daffalla, Tara Matthews, Sunny Consolvo, Jill Palzkill Woelfer, Patrick Gage Kelley, Michelle L. Mazurek, Dana Cuomo, Nicola Dell, and Thomas Ristenpart. Sok: Safer digital-safety research involving at-risk users. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 635–654, 2024. doi:10.1109/SP54263.2024.00071.

[50] Ruha Benjamin. Race after technology. In *Social Theory Re-Wired*, pages 405–415. Routledge, 2023.

[51] Betsy Woodruff Swan and Myah Ward. Trump's immigration crackdown is expected to start on Day 1. URL: https://www.politico.com/news/2024/11/18/immigration-100-days-trump-executive-action-00189286.

[52] David Bisson. Triada trojan conceals itself in whatsapp mod. https://web.archive.org/web/20250118125724/https://securityintelligence.com/news/triada-trojan-conceals-itself-whatsapp-mod/, October 2021.

[53] Kerstin Bongard-Blanchy, Ariana Rossi, Salvador Rivas, Sophie Doublet, Vincent Koening, and Gabriele Lenzini. "i am definitely manipulated, even when i am aware of it. it's ridiculous!" - dark patterns from the end-user perspective. In *Designing Interactive Systems Conference 2021*, 2021. URL: http://doi.org/10.1145/3461778.3462086.

[54] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. Tales from the dark side: Privacy dark strategies and privacy dark patterns.

2016(4):237–254, 2016. URL: https://content.sciendo.com/view/journals/popets/2016/4/article-p237.xml.

[55] Sarah Brayne and Angèle Christin. Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts. 68(3):608–624. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10798813/, arXiv:38250480, doi:10.1093/socpro/spaa004.

[56] Ryan Briggs. Krasner: "mass supervision" is the "evil twin" of mass incarceration. https://stoneleighfoundation.org/krasner-mass-supervision-is-the-evil-twin-of-mass-incarceration/, Mar 2019.

[57] Harry Brignull. Dark patterns. https://www.deceptive.design/types, 2010.

[58] Apr 2021. URL: https://www.ftc.gov/news-events/events-calendar/bringing-dark-patterns-light-ftc-workshop.

[59] Simone Browne. *Dark matters: On the surveillance of blackness*. Duke University Press, 2015.

[60] Amber M. Buck and Devon F. Ralston. I didn't sign up for your research study: The ethics of using "public" data. *Computers and Composition*, 61:102655, 2021. Rhetorics of Data: Collection, Consent, & Critical Digital Literacies. doi:10.1016/j.compcom.2021.102655.

[61] Joy Buolamwini and Timnit Gebru. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Sorelle A. Friedler and Christo Wilson, editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 77–91. PMLR. URL: https://proceedings.mlr.press/v81/buolamwini18a.html.

[62] John Burnett. Immigration Advocates Warn ICE Is Retaliating For Activism. URL: https://www.npr.org/2018/03/16/593884181/immigration-advocates-warn-ice-is-retaliating-for-activism.

[63] Kenneth P. Burnham and David R. Anderson. Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods & Research*, 33(2):261–304, 2004. arXiv:https://doi.org/10.1177/0049124104268644, doi:10.1177/0049124104268644.

[64] Kelly Caine. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 981–992. ACM. URL: https://dl.acm.org/doi/10.1145/2858036.2858498, doi:10.1145/2858036.2858498.

[65] Ryan Calo. Digital market manipulation. *George Washington Law Revision*, 82:995, 2013.

[66] Devon W Carbado, Kimberlé Williams Crenshaw, Vickie M Mays, and Barbara Tomlinson. Intersectionality: Mapping the movements of a theory1. *Du Bois review: social science research on race*, 10(2):303–312, 2013.

[67] Cassie Miller. Racist 'Replacement' Theory Believed by Half of Americans. URL: https://www.splcenter.org/resources/stories/poll-finds-support-great-replacement-hard-right-ideas/.

[68] Ismael Castell-Uroz, Xavier Marrugat-Plaza, Josep Solé-Pareta, and Pere Barlet-Ros. A first look into alexa's interaction security. In *Proceedings of the 15th International Conference on Emerging Networking EXperiments and Technologies*, CoNEXT '19 Companion, page 4–6, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3360468.3366769.

[69] Rahul Chatterjee, Periwinkle Doerfler, Hadas Orgad, Sam Havron, Jackeline Palmer, Diana Freed, Karen Levy, Nicola Dell, Damon McCoy, and Thomas Ristenpart. The Spyware Used in Intimate Partner Violence. In *Proc. IEEE S&P*, 2018.

[70] Christine Chen, Nicola Dell, and Franziska Roesner. Computer Security and Privacy in the Interactions Between Victim Service Providers and Human Trafficking Survivors. In *Proc. USENIX Security*, 2019.

[71] Kai Chen, Peng Wang, Yeonjoon Lee, XiaoFeng Wang, Nan Zhang, Heqing Huang, Wei Zou, and Peng Liu. Finding unknown malice in 10 seconds: mass vetting for new threats at the google-play scale. In *Proceedings of the 24th USENIX Conference on Security Symposium*, SEC'15, page 659–674, USA, 2015. USENIX Association.

[72] Nancy Chen. Asylum Employment Authorization, Explained. URL: `https://documentedny.com/2024/06/14/asylum-employment-authorization-document-clock/`.

[73] Long Cheng, Christin Wilson, Song Liao, Jeffrey Young, Daniel Dong, and Hongxin Hu. *Dangerous Skills Got Certified: Measuring the Trustworthiness of Skill Certification in Voice Personal Assistant Platforms*, page 1699–1716. Association for Computing Machinery, New York, NY, USA, 2020. URL: `https://doi.org/10.1145/3372297.3423339`.

[74] Shruthi Sai Chivukula, Chris Watkins, Lucca McKay, and Colin M. Gray. "nothing comes before profit": Asshole design in the wild. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–6, New York, NY, USA, 2019. Association for Computing Machinery. `doi:10.1145/3290607.3312863`.

[75] Christina Nuñez. 7 of the Biggest Challenges Immigrants and Refugees Face in the US. URL: `https://www.globalcitizen.org/en/content/the-7-biggest-challenges-facing-refugees-and-immig/`.

[76] Deborah Cohen and Benjamin Crabtree. Qualitative research guidelines project, Jul 2006. http://www.qualres.org/.

[77] Competition and Markets Authority of the United Kingdom. Evidence Review of Online Choice Architecture and Consumer and Competition Harm. Technical Report 157, April 2022. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1069423/OCA_Evidence_Review_Paper_14.4.22.pdf.

[78] The R User Conference. Technical note on cumulative link mixed models (clmms) in r with the package "ordinal", 2021. URL: https://user2021.r-project.org/participation/technical_notes/t186/technote//.

[79] Gregory Conti and Edward Sobiesk. Malicious interface design: Exploiting the user. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, page 271–280, New York, NY, USA, 2010. Association for Computing Machinery. doi:10.1145/1772690.1772719.

[80] Ronald P Corbett Jr. Burdens of leniency: The changing face of probation. *Minn. L. Rev.*, 99:1697, 2014.

[81] Chris Corde and Nimmy Reichenberg. Introducing the unified Chronicle Security Operations platform — Google Cloud Blog — cloud.google.com. https://cloud.google.com/blog/products/identity-security/introducing-the-unified-chronicle-security-operations-platform, September 2023.

[82] Corrisoft. Home. https://web.archive.org/web/20211012183449/https://corrisoft.com/, 2021.

[83] A Cortesi, M Hils, T Kriechbaumer, and contributors. mitmproxy. https://mitmproxy.org/, 2010.

[84] American Immigration Council. Asylum in the united states. https://www.americanimmigrationcouncil.org/sites/default/files/research/asylum_in_united_states_update_jan_2024.pdf, Jan 2024. [Accessed 12-12-2024].

[85] Robert M. Cover. Violence and the word. In Bruce B. Lawrence and Aisha Karim, editors, *A Reader*, pages 293–313. Duke University Press. URL: https://doi.org/10.1515/9780822390169-041, doi:doi:10.1515/9780822390169-041.

[86] Lorrie Faith Cranor, Adam L. Durity, Abigail Marsh, and Blase Ur. Parents' and teens' perspectives on privacy in a technology-filled world. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 19–35, Menlo Park, CA, July 2014. USENIX Association. URL: https://www.usenix.org/conference/soups2014/proceedings/presentation/cranor.

[87] Henry Dalziel and Ajin Abraham. *Automated security analysis of android and iOS applications with mobile security framework*. Syngress, 2015.

[88] Wes Davis. A woman and her daughter plead guilty to abortion-related charges supported by Meta-provided Facebook chats. URL: https://www.theverge.com/2023/7/11/23790923/facebook-meta-woman-daughter-guilty-abortion-nebraska-messenger-encryption-privacy.

[89] Google Developers. Detect when users start or end an activity. https://developer.android.com/guide/topics/location/transitions.

[90] Google Developers. Distribution dashboard. https://developer.android.com/about/dashboards.

[91] Google Developers. Manifest.permission. https://developer.android.com/reference/android/Manifest.permission.

[92] Google Developers. Request location permissions. https://developer.android.com/training/location/permissions.

[93] Shaila Dewan. Probation may sound light, but punishments can land hard. *The New York Times*, Aug 2015.

[94] Karel Dhondt, Victor Le Pochat, Yana Dimova, Wouter Joosen, and Stijn Volckaert. Swipe left for identity theft: An analysis of user data privacy risks on location-based dating apps. In *33rd USENIX Security Symposium (USENIX Security 24)*, pages 5053–5070, Philadelphia, PA, August 2024. USENIX Association. URL: https://www.usenix.org/conference/usenixsecurity24/presentation/dhondt.

[95] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. Ui dark patterns and where to find them: A study on mobile applications and user perception. In *Proceedings of the ACM on Human-Computer Interaction*, 2020.

[96] Cory Doctorow. African WhatsApp Modders are the Masters of Worldwide Adversarial Interoperability. https://www.eff.org/deeplinks/2020/03/african-whatsapp-modders-are-masters-worldwide-adversarial-interoperability, 2020.

[97] Daniel J. Dubois, Roman Kolcun, Anna Maria Mandalari, Muhammad Talha Paracha, David Choffnes, and Hamed Haddadi. When Speakers Are All Ears: Characterizing Misactivations of IoT Smart Speakers. In *Proc. of the Privacy Enhancing Technologies Symposium (PETS)*, 2020.

[98] dumbcollegekid. Alexa ignoring some basic voice commands, but responding to others., Mar 2017. URL: www.reddit.com/r/amazonecho/comments/5y5y1o/alexa_ignoring_some_basic_voice_commands_but/.

[99] Ingrid Eagly and Steven Shafer. Measuring in absentia removal in immigration court. *U. Pa. L. Rev.*, 168:817, 2019.

[100] Max Eddy. Rsac: Reverse-engineering an android app in five minutes. https://www.pcmag.com/news/rsac-reverse-engineering-an-android-app-in-five-minutes, February 2014.

[101] DIGIDAY Editors. Explainer: SDKs vs Libraries. https://digiday.com/media/explainer-sdks-vs-libraries/, Apr 2011.

[102] eHawk Solutions. Repath privacy policy. https://web.archive.org/web/20210927232836/https://ehawksolutions.com/repathprivacypolicy/.

[103] Elizabeth Warren. Warren, Cárdenas, Lawmakers Raise Concerns About Excessive Fees & Abusive Practices in the Electronic Monitoring and Private Probation Industries — U.S. Senator Elizabeth Warren of Massachusetts. URL: https://www.warren.senate.gov/newsroom/press-releases/warren-cardenas-lawmakers-raise-concerns-about-excessive-fees-and-abusive-practices-in-the-electronic-monitoring-and-private-probation-industries.

[104] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. Which privacy and security attributes most impact consumers' risk perception and willingness to purchase IoT devices? In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 1937–1954, 2021.

[105] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. Exploring how privacy and security factor into IoT device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[106] Enrico204. whatsapp-decoding/libwhatsapp at master · Enrico204/whatsapp-decoding — github.com. https://github.com/Enrico204/whatsapp-decoding/tree/master/libwhatsapp, 2017.

[107] Lauren Etter. What's the maker of post-it notes doing in the ankle monitor business? struggling. *Bloomberg.com*, Apr 2017.

[108] Virginia Eubanks. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.

[109] European Data Protection Board. Dark patterns in social media platform interfaces: How to recognise and avoid them, March 2022. https://edpb.europa.eu/system/files/2022-03/edpb_03-2022_guidelines_on_dark_patterns_in_social_media_platform_interfaces_en.pdf.

[110] Brian B Evans. Private prisons. *Emory LJ*, 36:253, 1987.

[111] Facebook. Permissions - facebook login - documentation. https://developers.facebook.com/docs/facebook-login/android/permissions/.

[112] Álvaro Feal, Paolo Calciati, Narseo Vallina-Rodriguez, Carmela Troncoso, Alessandra Gorla, et al. Angel or devil? a privacy study of mobile parental control apps. *Proceedings of Privacy Enhancing Technologies (PoPETS)*, 2020, 2020.

[113] Todd Feathers. 'They track every move': how US parole apps created digital prisoners. http://www.theguardian.com/global-development/2021/mar/04/they-track-every-move-how-us-parole-apps-created-digital-prisoners, Mar 2021.

[114] Chrome for Developers. https://developer.chrome.com/docs/crux/about/, June 2022.

[115] Directorate-General for Justice and Consumers (European Commission), Francisco Lupiáñez-Villanueva, Alba Boluda, Francesco Bogliacino, Giovanni Liva, Lucie Lechardoy, and Teresa Rodríguez de las Heras Ballell. Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation : final report. Technical report, Publications Office of the European Union, LU, 2022. URL: https://data.europa.eu/doi/10.2838/859030.

[116] The Wireshark Foundation. Wireshark. https://www.wireshark.org/.

[117] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology. In *Proc. CHI*, 2018.

[118] Alisa Frik, Leysan Nurgalieva, Julia Bernd, Joyce Lee, Florian Schaub, and Serge Egelman. Privacy and Security Threat Models and Mitigation Strategies of Older Adults. In *Proc. SOUPS*, 2019.

[119] FTC. A Brief Overview of the Federal Trade Commission's Investigative, Law Enforcement, and Rulemaking Authority — ftc.gov. `https://www.ftc.gov/about-ftc/mission/enforcement-authority`. [Accessed 09-04-2024].

[120] Sidney Fussell. Apps are now putting the parole agent in your pocket. `https://www.wired.com/story/apps-putting-parole-agent-your-pocket/`, Nov 2020.

[121] Robert S Gable. Left to Their Own Devices: Should Manufacturers of Offender Monitoring Equipment be Liable for Design Defect? *U. Ill. JL Tech. & Pol'y*, page 333, 2009.

[122] Benjamin Gardner. Incentivised snowballing. *The Psychologist*, 22(9):768–769, 2009.

[123] Timnit Gebru and Émile P Torres. The tescreal bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday*, 2024.

[124] Christine Geeng, Jevan Hutson, and Franziska Roesner. Usable sexurity: Studying People's concerns and strategies when sexting. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, pages 127–144. USENIX Association, August 2020. URL: `https://www.usenix.org/conference/soups2020/presentation/geeng`.

[125] Google. Device and network abuse. `https://support.google.com/googleplay/android-developer/answer/9888379#zippy=%2Cexamples-of-common-violations`.

[126] Google. Google play terms of service. `https://play.google.com/about/play-terms/index.html`, Oct 2020.

[127] Colin M. Gray, Jingle Chen, Shruthi Sai Chivukula, and Liyang Qu. End user accounts of dark patterns as felt manipulation. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):372:1–372:25, Oct 2021. doi:10.1145/3479516.

[128] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. The dark (patterns) side of ux design. In *Proceedings of the ACM on Human-Computer Interaction*, 2018.

[129] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. Dark patterns and the legal requirements of consent banners: An interaction criticism perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3411764.3445779.

[130] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. Dark and Bright Patterns in Cookie Consent Requests. *Journal of Digital Social Research*, 3(1):1–38, February 2021. Number: 1. URL: https://jdsr.se/ojs/index.php/jdsr/article/view/54, doi:10.33621/jdsr.v3i1.54.

[131] Ben Green. "Good" isn't good enough. In *AI for Social Good Workshop at NeurIPS*. URL: https://www.semanticscholar.org/paper/%E2%80%9CGood%E2%80%9D-isn%E2%80%99t-good-enough-Green/dc2fed36474b1d1dd497b8f08e06183bb65cf48f.

[132] Saira Hussain and Will Greenberg. Study of Electronic Monitoring Smartphone Apps Confirms Advocates' Concerns of Privacy Harms. URL: https://www.eff.org/deeplinks/2022/09/study-electronic-monitoring-smartphone-apps-confirms-advocates-concerns-privacy.

[133] Track Group. Track group privacy policy. https://web.archive.org/web/20210927153055/http://trackgrp.com/privacy-policy/, Jun 2020.

[134] Tamy Guberek, Allison McDonald, Sylvia Simioni, Abraham H. Mhaidli, Kentaro Toyama, and Florian Schaub. Keeping a Low Profile?: Technology, Risk and Privacy among Undocumented Immigrants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–15. ACM. URL: https://dl.acm.org/doi/10.1145/3173574.3173688, doi:10.1145/3173574.3173688.

[135] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. A comparative study of dark patterns across web and mobile modalities. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct 2021. doi:10.1145/3479521.

[136] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In *CHI Conference on Human Factors in Computing Systems*, pages 1–27, New Orleans LA USA, April 2022. ACM. URL: https://dl.acm.org/doi/10.1145/3491102.3501985, doi:10.1145/3491102.3501985.

[137] Catherine Han, Irwin Reyes, Álvaro Feal, Joel Reardon, Primal Wijesekera, Narseo Vallina-Rodriguez, Amit Elazari, Kenneth A. Bamberger, and Serge Egelman. The Price is (Not) Right: Comparing Privacy in Free and Paid Apps. *Proceedings on Privacy Enhancing Technologies*, 2020(3):222–242, Jul 2020. doi:10.2478/popets-2020-0050.

[138] Tom Hanson, Laura Geller, Josh Peña, John Kelly, Aaron Munoz, and Jose Sanchez. ICE's SmartLINK app tracks migrants by the thousands. Does it work? URL: https://www.cbsnews.com/news/does-ices-smartlink-app-work/.

[139] Woodrow Hartzog. The Case Against Idealising Control. 4(423). URL: https://papers.ssrn.com/abstract=3299762.

[140] Ayako A Hasegawa, Daisuke Inoue, and Mitsuaki Akiyama. A Survey on the Geographic Diversity of Usable Privacy and Security Research. *arXiv preprint arXiv:2305.05004*, 2023.

[141] Sam Havron, Diana Freed, Rahul Chatterjee, Damon McCoy, Nicola Dell, and Thomas Ristenpart. Clinical computer security for victims of intimate partner violence. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 105–122, Santa Clara, CA, August 2019. USENIX Association. URL: https://www.usenix.org/conference/usenixsecurity19/presentation/havron.

[142] Play Console Help. User data. https://support.google.com/googleplay/android-developer/answer/10144311.

[143] Joseph Henrich, Steven J Heine, and Ara Norenzayan. Most people are not WEIRD. *Nature*, 466(7302):29–29, 2010.

[144] Alexis Hiniker, Sungsoo (Ray) Hong, Tadayoshi Kohno, and Julie A. Kientz. Mytime: Designing and evaluating an intervention for smartphone non-use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 4746–4757, New York, NY, USA, 2016. Association for Computing Machinery. URL: https://homes.cs.washington.edu/~yoshi/papers/MyTimeCameraReady.pdf, doi:10.1145/2858036.2858403.

[145] Kelly Hoffman. *Use of Electronic Offender-Tracking Devices Expands Sharply.* Sep 2016.

[146] Andrew Horwitz. Coercion, pop-psychology, and judicial moralizing: Some proposals for curbing judicial abuse of probation conditions. *Wash. & Lee L. Rev.*, 57:75, 2000.

[147] Colin C. Ife, Yun Shen, Steven J. Murdoch, and Gianluca Stringhini. Marked for Disruption: Tracing the Evolution of Malware Delivery Operations Targeted for Takedown. In *Proc. RAID*, 2021.

[148] Muhammad Ikram, Narseo Vallina-Rodriguez, Suranga Seneviratne, Mohamed Ali Kaafar, and Vern Paxson. An analysis of the privacy and security risks of android

vpn permission-enabled apps. In *Proceedings of the 2016 internet measurement conference*, pages 349–364, 2016.

[149] U.S. Immigration and Customs Enforcement. Detention management—detention stastitics. https://www.ice.gov/detain/detention-management. Updated October 1, 2021.

[150] BI Incorporated. Bi smartlink™ privacy policy. https://web.archive.org/web/20210526234349/https://bi.com/products-and-services/bi-smartlink-privacy-policy/.

[151] IT Info. Root a device (rooting) – what does it mean? https://webllena.com/root-a-device-rooting-what-does-it-mean/, Feb 2018.

[152] Charles Koch Institute. What is community supervision? https://charleskochinstitute.org/stories/what-is-community-supervision/, Jan 2019.

[153] James Kilgore, Emmett Sanders, and Kate Weisburd. Carceral Surveillance and the Dangers of "Better-than-Incarceration" Reasoning. URL: https://lpeproject.org/blog/carceral-surveillance-and-the-dangers-of-better-than-incarceration-reasoning/.

[154] Jan and Brian. end-to-end encryption — blog.whatsapp.com. https://blog.whatsapp.com/end-to-end-encryption?lang=en_US, April 2016.

[155] Jeremy Hsu. Apps used as alternatives to prison in US found to have privacy flaws. URL: https://www.newscientist.com/article/2336473-apps-used-as-alternatives-to-prison-in-us-found-to-have-privacy-flaws/.

[156] Jimmie E. Gates and Alissa Zhu. Ankle monitors and informants: How ICE chose the 7 Mississippi food plants to raid. URL: https://www.clarionledger.com/story/news/politics/2019/08/09/ice-raids-federal-investigation-mississippi-poultry-plants/1960576001/.

[157] Deborah Johnson. Connections among poverty, incarceration, and inequality. https://www.irp.wisc.edu/resource/connections-among-poverty-incarceration-and-inequality/, May 2020.

[158] Alexi Jones. Correctional control 2018: Incarceration and supervision by state. https://www.prisonpolicy.org/reports/correctionalcontrol2018.html, Dec 2018.

[159] Jose Olivares and John Washington. ICE Discussed Punishing Immigrant Advocates for Peaceful Protests. URL: https://theintercept.com/2021/06/17/ice-retaliate-immigrant-advocates-surveillance/.

[160] Just Futures Law, Mijente Support Committee, and Community Justice Exchange. Fact Sheet on ICE FOIA Lawsuit: ICE Documents Reveal Alarming Scale of Surveillance in ISAP program. URL: https://static1.squarespace.com/static/62c3198c117dd661bd99eb3a/t/6512da273ccb7321c334ab6c/1695734312687/ATDFOIAFinal.pdf.

[161] Mannat Kaur, Michel van Eeten, Marijn Janssen, Kevin Borgolte, and Tobias Fiebig. Human Factors in Security Research: Lessons Learned from 2008-2018. *arXiv preprint arXiv:2103.13287*, 2021.

[162] Yomi Kazeem. WhatsApp is so popular in Africa, even knock-off versions are used more often than Facebook. https://qz.com/africa/1804859/fake-whatsapp-app-more-popular-than-facebook-instagram-in-africa, 2020.

[163] Sean Kennedy, Haipeng Li, Chenggang Wang, Hao Liu, Boyang Wang, and Wenhai Sun. I can hear your alexa: Voice command fingerprinting on smart home speakers. In *2019 IEEE Conference on Communications and Network Security (CNS)*, pages 232–240, 2019. doi:10.1109/CNS.2019.8802686.

[164] Os Keyes, Jevan Hutson, and Meredith Durbin. A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry.

In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11. ACM. URL: https://dl.acm.org/doi/10.1145/3290607.3310433, doi:10.1145/3290607.3310433.

[165] James Kilgore. Survey of EM Literature. https://www.challengingecarceration.org/survey-of-em-literature/, Jun 2018.

[166] James Kilgore. Let's fight for freedom from electronic monitors and e-carceration. https://truthout.org/articles/lets-fight-for-freedom-from-electronic-monitors-and-e-carceration/, Sep 2019.

[167] James Kilgore. As the u.s. scrambles to slow coronavirus, we should be wary of increased surveillance. https://theappeal.org/coronavirus-covid-19-surveillance-electronic-monitoring/, Mar 2020.

[168] Lauren Kilgour. The ethics of aesthetics: Stigma, information, and the politics of electronic ankle monitor design. *The Information Society*, 36(3):131–146, 2020.

[169] Jennifer King and Adriana Stephan. Regulating privacy dark patterns in practice - drawing inspiration from the california privacy rights act. *Georgetown Law Technology Review*, 5, 2021.

[170] Austin Kocher. Glitches in the Digitization of Asylum: How CBP One Turns Migrants' Smartphones into Mobile Borders. 13(6):149. URL: https://www.mdpi.com/2075-4698/13/6/149, doi:10.3390/soc13060149.

[171] Dan Komosny, Miroslav Voznak, and Saeed Ur Rehman. Location accuracy of commercial ip address geolocation databases. *Information Technology And Control*, 46(3):333–344, Sep 2017. doi:10.5755/j01.itc.46.3.14451.

[172] Veronika Krauss. Exploring dark patterns in xr. In *Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality, CHI Extended*

Abstracts, CHIEA '22. ACM, 2022. URL: https://wenjietseng.com/assets/pdf/SSPXR22_submissions/SSPXR22_paper_8.pdf.

[173] Chiara Krisam, Heike Dietmann, Melanie Volkamer, and Oksana Kulyk. Dark patterns in the wild: Review of cookie disclaimer designs on top 500 german websites. In *European Symposium on Usable Security 2021*, EuroUSEC '21, page 1–8, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3481357.3481516.

[174] Cherie Lacey and Catherine Caudwell. Cuteness as a 'dark pattern' in home robots. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, HRI '19, page 374–381. IEEE Press, 2019.

[175] Langston Hughes. Mother to Son.

[176] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), nov 2018. doi:10.1145/3274371.

[177] Ada Lerner, Helen Yuxun He, Anna Kawakami, Silvia Catherine Zeamer, and Roberto Hoyle. Privacy and activism in the transgender community. In *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–13, Honolulu, HI, U.S.A., Apr 2020. ACM. URL: https://dl.acm.org/doi/10.1145/3313831.3376339, doi:10.1145/3313831.3376339.

[178] James Andrew Lewis. Technology and the Shifting Balance of Power, Tue, 04/19/2022 - 12:00. URL: https://www.csis.org/analysis/technology-and-shifting-balance-power.

[179] Li Li, Tegawendé F. Bissyandé, and Jacques Klein. Rebooting research on detecting repackaged android apps: Literature review and benchmark. *IEEE Transactions on Software Engineering*, 47(4):676–693, 2021. doi:10.1109/TSE.2019.2901679.

[180] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. How weird is chi? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3411764.3445488.

[181] Outreach Smartphone Monitoring LLC. Privacy policy. https://web.archive.org/web/20210929230231/https://app.osmnow.com/privacy_policy.pdf, Feb 2020.

[182] TRACKtech LLC. Privacy policy. https://web.archive.org/web/20210921161420/https://app.tracktechllc.com/privacy_policy.html.

[183] Peter Lowe. Peter lowe's ad and tracking server list. https://pgl.yoyo.org/adservers/serverlist.php?hostformat=hosts&showintro=1&mimetype=plaintext, January 2024.

[184] Jamie Luguri and Lior Strahilevitz. Shining a light on dark patterns. *U of Chicago, Public Law Working Paper No. 719; University of Chicago Coase-Sandor Institute for Law & Economics Research Paper No. 879*, 2019. URL: https://ssrn.com/abstract=3431205.

[185] Aditi M. Bhoot, Mayuri A. Shinde, and Wricha P. Mishra. Towards the identification of dark patterns: An analysis based on end-user reactions. In *IndiaHCI '20: Proceedings of the 11th Indian Conference on Human-Computer Interaction*, IndiaHCI 2020, page 24–33, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3429290.3429293.

[186] Eryn Ma and Eleanor Birrell. Prospective Consent: The Effect of Framing on Cookie Consent Decisions. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, CHI EA '22, pages 1–6, New York, NY, USA, April 2022. Association for Computing Machinery. doi:10.1145/3491101.3519687.

[187] Xiao Ma and Ariel Liu. Challenges in supporting exploratory search through voice assistants. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, CUI '20, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3405755.3406152.

[188] Mary Madden. Privacy, security, and digital inequality. https://datasociety.net/wp-content/uploads/2017/09/DataAndSociety_PrivacySecurityandDigitalInequality.pdf, September 2017.

[189] Matt Mahmoudi. Introduction. In *Migrants in the Digital Periphery*, New Urban Frontiers of Control, pages 1–20. University of California Press, 1 edition. URL: https://www.jstor.org/stable/jj.25494964.6, arXiv:jj.25494964.6, doi:10.2307/jj.25494964.6.

[190] David Major, Danny Yuxing Huang, Marshini Chetty, and Nick Feamster. Alexa, who am i speaking to?: Understanding users' ability to identify third-party apps on amazon alexa. *ACM Trans. Internet Technol.*, 22(1), sep 2021. doi:10.1145/3446389.

[191] Marc Cieslak. DeepSeek shows AI's centre of power could shift away from US. URL: https://www.bbc.com/news/articles/c9w5d9new0yo.

[192] Shrirang Mare, Logan Girvin, Franziska Roesner, and Tadayoshi Kohno. Consumer smart homes: Where we are and where we need to go. In *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications*, HotMobile '19, page 117–122, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3301293.3302371.

[193] Sergio Martínez-Beltrán. President Trump's suspension of asylum marks a break from U.S. past. URL: https://www.npr.org/2025/01/23/nx-s1-5272406/trump-suspends-asylum.

[194] Silvia Masiero. Should we still be doing ICT4D research? 88(5):e12215. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/isd2.12215, doi:10.1002/isd2.12215.

[195] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. Dark patterns at scale: Findings from a crawl of 11k shopping websites. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019. doi:10.1145/3359183.

[196] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3411764.3445610.

[197] MaxMind. Geoip2 precision web service demo. https://www.maxmind.com/en/geoip2-precision-demo.

[198] MAXQDA. https://www.maxqda.com/, 2024.

[199] Allison McDonald, Catherine Barwulor, Michelle L. Mazurek, Florian Schaub, and Elissa M. Redmiles. "It's stressful having all these phones": Investigating Sex Workers' Safety Goals, Risks, and Practices Online. In *Proc. USENIX Security*, 2021.

[200] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Human-Computer Interaction*, 3(CSCW), 2019.

[201] Tomas Meskauskas. Triada trojan (android). https://www.pcrisk.com/removal-guides/24926-triada-trojan-android, November 2023.

[202] Gillian E Metzger. Privatization as delegation. *Colum. L. Rev.*, 103:1367, 2003.

[203] Abraham Hani Mhaidli and Florian Schaub. Identifying manipulative advertising techniques in xr through scenario construction. In *CHI Conference on Human Factors in Computing Systems*, 2021.

[204] Just Futures Law & Mijente. Ice digital prisons. `https://www.flipsnack.com/JustFutures/ice-digital-prisons-1u8w3fnd1j.html`, May 2021.

[205] Thomas Mildner and Gian-Luca Savino. *Ethical User Interfaces: Exploring the Effects of Dark Patterns on Facebook*. Association for Computing Machinery, New York, NY, USA, 2021. URL: `https://doi.org/10.1145/3411763.3451659`.

[206] Alberto Monge Roffarello and Luigi De Russis. Towards Understanding the Dark Patterns That Steal Our Attention. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, CHI EA '22, pages 1–7, New York, NY, USA, April 2022. Association for Computing Machinery. `doi:10.1145/3491101.3519829`.

[207] Camilo Montoya-Galvez. ICE to test smartwatch-like tracking devices for migrants facing deportation. URL: `https://www.cbsnews.com/news/ice-smartwatch-tracking-device-migrants-deportation/`.

[208] Collins W. Munyendo, Yasemin Acar, and Adam J. Aviv. "Desperate Times Call for Desperate Measures": User Concerns with Mobile Loan Apps in Kenya. In *Proc. IEEE S&P*, 2022.

[209] Collins W. Munyendo, Yasemin Acar, and Adam J. Aviv. "In Eighty Percent of the Cases, I Select the Password for Them": Security and Privacy Challenges, Advice, and Opportunities at Cybercafes in Kenya. In *Proc. IEEE S&P*, 2023.

[210] Collins W. Munyendo, Kentrell Owens, Faith Strong, Shaoqi Wang, Adam J. Aviv, Tadayoshi Kohno, and Franziska Roesner. "You Have to Ignore the Dangers": User Perceptions of the Security and Privacy Benefits of WhatsApp Mods . In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 4515–4533, Los Alamitos, CA, USA,

May 2025. IEEE Computer Society. URL: https://doi.ieeecomputersociety.org/10.1109/SP61157.2025.00087, doi:10.1109/SP61157.2025.00087.

[211] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy expectations and preferences in an IoT world. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 399–412, 2017.

[212] Sheza Naveed, Hamza Naveed, Mobin Javed, and Maryam Mustafa. "Ask this from the person who has private stuff": Privacy Perceptions, Behaviours and Beliefs Beyond W.E.I.R.D. In *CHI Conference on Human Factors in Computing Systems*, pages 1–17. ACM. URL: https://dl.acm.org/doi/10.1145/3491102.3501883, doi:10.1145/3491102.3501883.

[213] Nedah Nemati and Dasha Pruss. Carceral technology and the normalization of psychological torture. URL: https://www.openglobalrights.org/carceral-technology-and-the-normalization-of-psychological-torture/.

[214] Mike Nellis. "Better than Human"? Smartphones, Artificial Intelligence and Ultra-Punitive Electronic Monitoring. *Challenging E-Carceration*, page 20, 2019.

[215] Trung Tin Nguyen, Michael Backes, Ninja Marnau, and Ben Stock. Share First, Ask Later (or Never?) Studying Violations of GDPR's Explicit Consent in Android Apps. In *30th USENIX Security Symposium*, August 2021.

[216] Yazmine Nichols. For Defense Attorneys: Tips for Effectively Challenging Pretrial Electronic Monitoring — ACLU. URL: https://www.aclu.org/news/criminal-law-reform/defense-attorneys-tips-for-challenging-electronic-monitoring.

[217] D. Nuñez. Inside the Border, Outside the Law: Undocumented Immigrants and the Fourth Amendment. 85:85–139. URL: https://digitalcommons.law.byu.edu/faculty_scholarship/108.

[218] Nathan Oyori Ogechi. On language rights in Kenya. *Nordic Journal of African Studies*, 12(3):19–19, 2003.

[219] National Commission on Informatics and Liberty. *Shaping Choices in the Digital World*. 2020. URL: https://linc.cnil.fr/sites/default/files/atoms/files/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf.

[220] OpenAI. Whisper. URL: https://github.com/openai/whisper.

[221] Molly Osberg and Dhruv Mehrotra. When your freedom depends on an app. https://gizmodo.com/when-your-freedom-depends-on-an-app-1843109198, Apr 2020.

[222] Kentrell Owens, Anita Alem, Franziska Roesner, and Tadayoshi Kohno. Electronic monitoring smartphone apps: An analysis of risks from technical, Human-Centered, and legal perspectives. In *31st USENIX Security Symposium (USENIX Security 22)*, Boston, MA, August 2022. USENIX Association. URL: https://www.usenix.org/conference/usenixsecurity22/presentation/owens.

[223] Kentrell Owens, Camille Cobb, and Lorrie Cranor. "You Gotta Watch What You Say": Surveillance of Communication with Incarcerated People. In *CHI '21*, May 2021. doi:10.1145/3411764.3445055.

[224] Kentrell Owens, Yael Eiger, Basia Radka, Tadayoshi Kohno, and Franziska Roesner. Understanding experiences with compulsory immigration surveillance in the u.s. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*, page 13, New York, NY, USA, June 2025. ACM. doi:10.1145/3715275.3732057.

[225] Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner. Exploring deceptive design patterns in voice interfaces. In *2022 European Symposium on Usable Security*, EuroUSEC 2022, page 64–78, New

York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3549015.3554213.

[226] PaymentsJournal. Prepaid card trends by age and income:. https://www.paymentsjournal.com/prepaid-card-trends-by-age-and-income/, Nov 2020.

[227] Alan Peshkin. The goodness of qualitative research. *Educational researcher*, 22(2):23–29, 1993.

[228] Mark Pogrebin, Mary Dodge, and Paul Katsampes. The collateral costs of short-term jail incarceration: The long-term social and economic disruptions. *Corrections Management Quarterly*, 5:64–69, 2001.

[229] Alisha Pradhan, Kanika Mehta, and Leah Findlater. "accessibility came by accident": Use of voice-controlled intelligent personal assistants by people with disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 1–13, Montreal QC Canada, Apr 2018. ACM. URL: https://dl.acm.org/doi/10.1145/3173574.3174033, doi:10.1145/3173574.3174033.

[230] The Associated Press. Ring will no longer allow police to request users' doorbell camera footage. URL: https://www.npr.org/2024/01/25/1226942087/ring-will-no-longer-allow-police-to-request-users-doorbell-camera-footage.

[231] Exodus Privacy. εxodus: The privacy audit platform for android applications. https://reports.exodus-privacy.eu.org/en/.

[232] Dasha Pruss, Hannah Pullen-Blasnik, Nikki Stevens, Shakeer Rahman, Clara Belitz, Logan Stapleton, Mallika G. Dharmaraj, Mizue Aizeki, Petra Molnar, Annika Pinch, Nathan Ryan, Thallita Lima, David Gray Widder, Amiya Tiwari, Ly Xīnzhèn Zhǎngsūn Brown, Jason S. Sexton, and Pablo Nunes. Prediction and Punishment: Critical Report on Carceral AI. URL: https://www.ssrn.com/abstract=5017321, doi:10.2139/ssrn.5017321.

[233] Alok Rai. #20 Amazing Android Modded Apps to use in 2023 - Today's Tech World — todaystechworld.com. https://todaystechworld.com/top-android-modded-apps/, January 2023.

[234] Abbas Razaghpanah, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Christian Kreibich, Phillipa Gill, Mark Allman, and Vern Paxson. Haystack: In situ mobile traffic analysis in user space. *arXiv preprint arXiv:1510.01419*, pages 1–13, 2015.

[235] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. Where is the digital divide? a survey of security, privacy, and socioeconomics. In *Proc. of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 931–936, New York, NY, USA, 2017. Association for Computing Machinery. doi:10.1145/3025453.3025673.

[236] Jake Reichel, Fleming Peck, Mikako Inaba, Bisrat Moges, Brahmnoor Singh Chawla, and Marshini Chetty. 'I have too much respect for my elders': Understanding South African Mobile Users' Perceptions of Privacy and Current Behaviors on Facebook and WhatsApp. In *Proc. USENIX Security*, 2020.

[237] Jingjing Ren, Daniel J. Dubois, David Choffnes, Anna Maria Mandalari, Roman Kolcun, and Hamed Haddadi. Information exposure from consumer iot devices: A multidimensional, network-informed measurement approach. In *Proceedings of the Internet Measurement Conference*, IMC '19, page 267–279, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3355369.3355577.

[238] Neil M Richards. THE DANGERS OF SURVEILLANCE. 126:32.

[239] Pedro Domínguez Rojas. That cool whatsapp mod you've installed comes with a surprise, and not a good one. https://en.softonic.com/articles/that-whatsapp-mod-so-cool-that-you-installed-comes-with-a-surprise-and-not-the-good-kind, 2023.

[240] Joel Rose. Immigrant Activists Say ICE Is Purposely Targeting Them. They're Urging Biden To Help. URL: https://www.npr.org/2021/08/04/1024348198/immigrant-activists-ask-biden-administration-to-ban-ice-from-retaliating-against.

[241] Jed Rubenfeld. Privatization and state action: Do campus sexual assault hearings violate due process. *Tex. L. Rev.*, 96:15, 2017.

[242] Luis A. Saavedra, Hridoy S. Dutta, Alastair R. Beresford, and Alice Hutchings. ModZoo: A Large-Scale Study of Modded Android Apps and their Markets. In *2024 APWG Symposium on Electronic Crime Research (eCrime)*, pages 162–174. URL: http://arxiv.org/abs/2402.19180, arXiv:2402.19180, doi:10.1109/eCrime66200.2024.00018.

[243] Aafaq Sabir, Evan Lafontaine, and Anupam Das. Hey alexa, who am i talking to?: Analyzing users' perception and awareness regarding third-party alexa skills. In *CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3491102.3517510.

[244] Georgia Robins Sadler, Hau-Chen Lee, Rod Seung-Hwan Lim, and Judith Fullerton. Research article: Recruitment of hard-to-reach population subgroups via adaptations of the snowball sampling strategy. *Nursing & Health Sciences*, 12(3):369–374, 2010. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1442-2018.2010.00541.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1442-2018.2010.00541.x, doi:10.1111/j.1442-2018.2010.00541.x.

[245] Johnny Saldaña. *The coding manual for qualitative researchers*. sage, 2009.

[246] Jerome H Saltzer and Michael D Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.

[247] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchill.

"Privacy is not for me, it's for those rich women": Performative Privacy Practices on Mobile Phones by Women in South Asia. In *Proc. SOUPS*, 2018.

[248] Saneta deVuono-powell, Chris Schweidler, Alicia Walters, and Azadeh Zohrabi. Who Pays? The True Cost of Incarceration on Families. URL: https://ellabakercenter.org/wp-content/uploads/2022/09/Who-Pays-FINAL.pdf.

[249] Wendy Sawyer and Peter Wagner. Mass incarceration: The whole pie 2020. https://www.prisonpolicy.org/reports/pie2020.html, Mar 2020.

[250] Vincent Schiraldi. Explainer: How 'technical violations' drive incarceration. https://theappeal.org/the-lab/explainers/explainer-how-technical-violations-drive-incarceration/, March 2021.

[251] R Schulkind, W Brade, J Hynes, and K Allinson. Every move you make: The human cost of gps tagging in the immigration system. *Report, Bail for Immigration Detainees, Medical Justice, and the Public Law Project, UK*, 2022.

[252] Adam Schwartz. Digital Privacy at the U.S. Border: Protecting the Data On Your Devices. URL: https://www.eff.org/wp/digital-privacy-us-border-2017.

[253] Andy J. Semotiuk. Constitutional Rights Of Undocumented Immigrants: Do They Have Any? URL: https://www.forbes.com/sites/andyjsemotiuk/2025/05/06/constitutional-rights-of--undocumented-immigrants-do-they-have-any/.

[254] Ari Shapiro. Months After Massive ICE Raid, Residents Of A Mississippi Town Wait And Worry. URL: https://www.npr.org/2019/11/17/778611834/months-after-massive-ice-raid-residents-of-a-mississippi-town-wait-and-worry.

[255] Sarah Sherman-Stokes. Immigration Detention Abolition and the Violence of Digital Cages. 95(1):219–266. URL: https://heinonline.org/HOL/P?h=hein.journals/ucollr95&i=233.

[256] Lukasz Siewierski and Android Security & Privacy Team. Pha family highlights: Triada, June 2019. URL: https://security.googleblog.com/2019/06/pha-family-highlights-triada.html.

[257] Mirela Silva and Daniela Oliveira. Brazilian favela women: How your standard solutions for technology abuse might actually harm them, Aug 2020.

[258] Lucy Simko, Ada Lerner, Samia Ibtasam, Franziska Roesner, and Tadayoshi Kohno. Computer Security and Privacy for Refugees in the United States. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 409–423. IEEE. URL: https://ieeexplore.ieee.org/document/8418616/, doi:10.1109/SP.2018.00023.

[259] Caroline Sinders. What's in a name? unpacking dark patterns versus deceptive design, June 2022. https://medium.com/@carolinesinders/whats-in-a-name-unpacking-dark-patterns-versus-deceptive-design-e96068627ec4.

[260] Preston So. Affordance and wayfinding in voice interface design, Mar 2020. URL: https://preston.so/writing/affordance-and-wayfinding-in-voice-interface-design/.

[261] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. *Circumvention by Design - Dark Patterns in Cookie Consent for Online News Outlets*. Association for Computing Machinery, New York, NY, USA, 2020. URL: https://doi.org/10.1145/3419249.3420132.

[262] Statista. Most popular messaging apps 2024 — Statista — statista.com. https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/, 2024.

[263] Kristen Stephens. The role of ethics in voice assistant design, Oct 2021. URL: https://voices.soundhound.com/the-role-of-ethics-in-voice-assistant-design/.

[264] Guillermo Suarez-Tangil and Gianluca Stringhini. Eight years of rider measurement in the android malware ecosystem. *IEEE Transactions on Dependable and Secure Computing*, 19(1):107–118, 2022. doi:10.1109/TDSC.2020.2982635.

[265] Cameron Summerson. How to sideload apps on android. https://www.howtogeek.com/313433/how-to-sideload-apps-on-android/, Jan 2018.

[266] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. Investigating users' preferences and expectations for always-listening voice assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(4), dec 2019. doi:10.1145/3369807.

[267] Diane Taylor. GPS tagging migrants 'psychological torture', says report. URL: https://www.theguardian.com/uk-news/2022/oct/31/gps-tagging-migrants-psychological-torture-says-report.

[268] Shadowtrack Technologies. Privacy policy. https://web.archive.org/web/20210528161904/https://www.shadowtrack.com/about_us/privacy/.

[269] Telmate. Telmate guardian: A new frontier in smartphone-based location monitoring; the case for telmate guardian and "community corrections," including parole, probation, pre-trial & work release. https://web.archive.org/web/20210511011316/https://www.telmate.com/wp-content/uploads/2015/07/TEL-WhitePaper-Guardian-NonBleed-r6.pdf, Jul 2015.

[270] The Bail Project. Beyond Bail: A National Survey of Pretrial Justice Reform in the United States. URL: https://www.bailproject.org/beyond-bail.

[271] Matthew Tokson. When the Government Buys Sensitive Personal Data. 11/3/2023 1:47:23 PM. URL: https://www.lawfaremedia.org/article/when-the-government-buys-sensitive-personal-data.

[272] Kentaro Toyama. Technology as amplifier in international development. In *Proceedings of the 2011 iConference*, pages 75–82. 2011.

[273] TRAC. TRAC Immigration - Comprehensive, independent, and nonpartisan information about immigration enforcement - alternatives to detention (ATD). URL: https://tracreports.org/immigration/detentionstats/atd_pop_table.html.

[274] Elizabeth Trovall. The growing business of immigrant surveillance. URL: https://www.marketplace.org/2023/08/02/the-growing-business-of-immigrant-surveillance/.

[275] U. S. Government Accountability Office. Alternatives to Detention: ICE Needs to Better Assess Program Performance and Improve Contract Oversight — U.S. GAO. URL: https://www.gao.gov/products/gao-22-104529.

[276] U.S. Immigration and Customs Enforcement. Alternatives to Detention Frequently Asked Questions. URL: https://www.ice.gov/atd-faq.

[277] Narseo Vallina-Rodriguez. Lumen privacy monitor. https://www.icsi.berkeley.edu/icsi/projects/networking/haystack.

[278] Alicia Virani. *Pretrial Electronic Monitoring in Los Angeles County.* Feb 2022.

[279] Jessica Vitak, Michael Zimmer, Anna Lenhart, Sunyup Park, Richmond Y. Wong, and Yaxing Yao. Designing for Data Awareness: Addressing Privacy and Security Concerns About "Smart" Technologies. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, pages 364–367, Virtual Event USA, October 2021. ACM. URL: https://dl.acm.org/doi/10.1145/3462204.3481724, doi:10.1145/3462204.3481724.

[280] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. Eliciting and analysing users' envisioned dialogues with perfect voice

assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3411764.3445536.

[281] Haoyu Wang, Zhe Liu, Jingyue Liang, Narseo Vallina-Rodriguez, Yao Guo, Li Li, Juan Tapiador, Jingcun Cao, and Guoai Xu. Beyond Google Play: A Large-Scale Comparative Study of Chinese Android App Markets. In *Proc. IMC*, 2018.

[282] Mark Warschauer, Michele Knobel, and Leeann Stone. Technology and Equity in Schooling: Deconstructing the Digital Divide. 18(4):562–588. URL: https://journals.sagepub.com/doi/10.1177/0895904804266469, doi:10.1177/0895904804266469.

[283] Kate Weisburd. Sentenced to surveillance: Fourth amendment limits on electronic monitoring. *NCL Rev.*, 98:717, 2019.

[284] Kate Weisburd. Punitive surveillance. *Virginia Law Review*, 108, 2021.

[285] Kate Weisburd, Varun Bhadha, Matthew Clauson, Jeanmarie Elican, Fatima Kahn, Kendall Lawrenz, Brooke Pemberton, Rebecca Ringler, Jordan Schaer, Mikayla Sherman, and Sarah Wohlsdorf. Electronic prisons: The operation of ankle monitoring in the criminal legal system. *George Washington University Law School*, page 54, 2021.

[286] WhatsApp. Two Billion Users – Connecting the World Privately — blog.whatsapp.com. https://blog.whatsapp.com/two-billion-users-connecting-the-world-privately, Feb 2020.

[287] WhatsApp. New Features for More Privacy, More Protection, More Control — blog.whatsapp.com. https://blog.whatsapp.com/new-features-for-more-privacy-more-protection-more-control, August 2022.

[288] WhatsApp. Chat Lock: Making your most intimate conversations even more private — blog.whatsapp.com. https://blog.whatsapp.com/chat-lock-making-your-most-intimate-conversations-even-more-private, May 2023.

[289] WhatsApp. Introducing Secret Code for Chat Lock — blog.whatsapp.com. https://blog.whatsapp.com/introducing-secret-code-for-chat-lock, November 2023.

[290] WhatsApp. Multiple Accounts Coming to WhatsApp — blog.whatsapp.com. https://blog.whatsapp.com/multiple-accounts-coming-to-whatsapp, October 2023.

[291] WhatsApp. About disappearing messages — WhatsApp Help Center — faq.whatsapp.com. https://faq.whatsapp.com/673193694148537, 2024.

[292] WhatsApp. How to check read receipts. https://faq.whatsapp.com/665923838265756/?cms_platform=android, 2024.

[293] WhatsApp. How to silence unknown callers — WhatsApp Help Center — faq.whatsapp.com. https://faq.whatsapp.com/1238612517047244, 2024.

[294] GB WhatsApp. GBWhatsApp Pro APK V17.60 Download For Android (2024) — gbwhatspro.com. https://gbwhatspro.com/, 2024.

[295] Lindsay Whitehurst and Gene Johnson. Appeals court allows Trump administration to suspend approval of new refugees amid lawsuit. URL: https://apnews.com/article/refugee-program-trump-administration-appeals-court-a6188722de3e3e1d2f344862b853d0c7.

[296] Zack Whittaker. 'Reverse' searches: The sneaky ways that police tap tech companies for your private data. URL: https://techcrunch.com/2024/04/02/reverse-searches-police-tap-tech-companies-private-data/.

[297] Langdon Winner. Do Artifacts Have Politics? 109(1):121–136. URL: https://www.jstor.org/stable/20024652, arXiv:20024652.

[298] Sally Wyatt. Technological determinism is dead; long live technological determinism. *The handbook of science and technology studies*, 3:165–180, 2008.

[299] Yaxing Yao, Richmond Wong, Pardis Emami-Naeini, Nick Merrill, Xinru Page, Yang Wang, and Pamela Wisniewski. Ubiquitous privacy: Research and design for mobile and iot platforms. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, CSCW '19, page 533–538, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3311957.3359430.

[300] Daniel Yeager. Certain certiorari: The digital privacy rights of probationers. *Conn. L. Rev. CONNtemplations*, 50:1, 2017.

[301] Min Zheng, Patrick PC Lee, and John CS Lui. Adam: an automatic and extensible platform to stress test android anti-virus systems. In *International conference on detection of intrusions and malware, and vulnerability assessment*, pages 82–101. Springer, 2012.

[302] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. User perceptions of smart home iot privacy. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), nov 2018. doi:10.1145/3274469.

[303] Wu Zhou, Yajin Zhou, Xuxian Jiang, and Peng Ning. Detecting repackaged smartphone applications in third-party android marketplaces. In *Proc. CODASPY*, 2012.

[304] Yajin Zhou and Xuxian Jiang. Dissecting android malware: Characterization and evolution. In *2012 IEEE Symposium on Security and Privacy*, pages 95–109, 2012. doi:10.1109/SP.2012.16.

[305] Catherine Zhu. Dark patterns — a new frontier in privacy regulation, July 2021. https://www.reuters.com/legal/legalindustry/dark-patterns-new-frontier-privacy-regulation-2021-07-29/.

[306] Shuofei Zhu, Ziyi Zhang, Limin Yang, Linhai Song, and Gang Wang. Benchmarking Label Dynamics of VirusTotal Engines. In *Proc. CCS*, 2020.

Appendix A

ADDITIONAL MATERIAL FOR ELECTRONIC MONITORING SMARTPHONE APPS

A.1 Review qualitative analysis codes

- Sentiment: Positive, with explanation
- Sentiment: Positive, no explanation
- Sentiment: Negative, proper function
- Sentiment: Negative, with malfunction
- Sentiment: Negative, no explanation or unclear explanation
- Sentiment: Neutral or unclear
- Justice: Felt wronged, injustice
- Sensors: Problem with camera
- Sensors: Problem with microphone
- Sensors: Problem with location
- Sensors: Problem with external device
- Authentication: Unable to login
- Malfunctions: Problems after update

- Malfunctions: Check-in or monitoring features not working
- Malfunctions: Faulty notification behavior
- Risks: Getting a violation because of faulty app
- Risks: Surveillance/privacy
- Comparison to alternatives: Better than traditional alternatives
- Comparison to alternatives: Worse than traditional alternatives
- Comparison to alternatives: Better than Prison
- Disruptions: Device limitations
- Disruptions: Loud alerts
- Misc: Technical Support Requests/Issues
- Misc: Took screenshots to capture check-ins that were logged
- Misc: Billing issues
- Misc: Forced to remove mask to check-in

A.2 Privacy policy qualitative analysis codes

- Mentions their mobile app?
- Mentions Mobile Data collection? E.g., location, contacts, camera
- Mentions sharing with law enforcement (without warrant)?

- Mentions sharing with 3rd parties?
- Mention any regulations?
- Mentions processing data on servers worldwide?
- Software updates mentioned?
- Do they sell your data?
- Do they mention "marketing purposes"?
- Mentions Retention? / Deletion?

A.2.1 Summary of Network Traffic Analysis

App Name	Potentially Identifying Domain(s) Contacted	Third Party Libraries
`Telmate Guardian`	api.telmateguardian.com	Flurry, New Relic, Urban Airship
`BI SmartLINK`	{bicdn, services, services.tn}.bi.com	Google Firebase Analytics, Microsoft Visual Studio App Center Analytics
`RePath`	app-version-log.repathportal.com	Google CrashLytics, Google Firebase Analytics
`IntelliTrack Mobile`	intellitrack-api.trackgrp.com	Google CrashLytics, Google Firebase Analytics
`Community Supervision`	api.globalsupervision.net	Xamarin
`aCheck`	gwusacheck.aware.attentigroup.com	
`Reconnect Community`		Google Firebase Analytics, Microsoft Visual Studio App Center Analytics, Microsoft Visual Studio App Center Crashes
`Uptrust`		Facebook, Google Firebase Analytics
`Outreach Smartphone Monitoring`		Google CrashLytics, Google Firebase Analytics
`Shadowtrack`		Google CrashLytics, Google Firebase Analytics
`TRACKphone Lite`		Branch
`Sprokit`		Facebook

Table A.1: Summary of network traffic analysis.

A.2.2 App Information

App Name	App ID	Installs	Additional usage information
`aCheck`	com.attenti.acheck.us	100+	
`BI SmartLINK`	com.biinc.mobile.client	100,000+	In 2020, the `BI SmartLINK` app was used by Immigration & Customs Enforcement (ICE) to monitor approx. 24,000 people; as of May 8, 2021, this number is 34,445 people [113, 149].
`Community Supervision`	com.supervision.community	500+	
`Corrisoft AIR Check-In`	com.corrisoft.air.core	1,000+	
`Securus Enrollink`	com.stopllc.offendermobile	1,000+	
`IntelliTrack Mobile`	com.trackgrp.intellitrackmobile	100+	
`Omnilink FocalPoint`	com.numerex.focalpoint	1,000+	
`Outreach Smartphone Monitoring`	com.osmnow	1,000+	
`Reconnect Community`	org.call2test.connectcomply	10,000+	
`RePath`	com.ehawk.repath	1,000+	
`Shadowtrack`	com.shadowtrack.shadowtrackview	10,000+	The `Shadowtrack` app is being used by approx. 11,000 people on probation in Virginia [113].
`Sprokit`	com.sprokit.Sprokit	100+	No longer available as of February 2022.
`TBV`	com.tbv.totalrecovery	100+	
`Telmate Guardian`	com.telmate.prod	10,000+	
`TRACKphone Lite`	com.tracktechllc.trackphonelite	100+	
`Uptrust`	com.uptrust.enduser	100+	

Table A.2: EM app details, including Google Play Store installs and additional usage information.

Appendix B

ADDITIONAL MATERIAL FOR UNDERSTANDING IMMIGRATION SURVEILLANCE

B.1 Interview protocol

Below is the interview protocol we used during our semi-structured interviews. The semi-structured nature of these interviews meant that we did not ask every question below in every interview, and all questions that we asked are not captured below. For example, whenever a participant mentioned a topic with which we were unfamiliar, we asked several clarifying questions to solidify our understanding of the topic before moving on. Naturally, these type of clarifying questions were more prevalent in earlier interviews and became less prevalent over time.

Consent form

Hello, thank you for agreeing to participate in this user study. Did you have a chance to look at the consent form before? No worries if not; we will go through it right now.

Rapport building

Q1 What type of work do you do as an advocate?

Q2 How many directly-impacted people have you worked with in the past?

Q3 How did you become an immigrant rights advocate? What motivates you to do this work?

Q4 What are the typical backgrounds of people with whom you work?

Main study

One goal of this research is to understand what questions people have about this app and how it functions. In future work we would like to attempt to answer these questions. For

today we cannot offer answers to these questions because we likely do not yet know them but we would like to hear what they are. We would like to begin this interview by asking you to think about what questions you have about how the app functions or what it does. You can voice these questions now or you can voice the questions as they come to you during the rest of the interview. It may be the case that discussing a specific topic might remind you of questions that you have heard in the past. There are no stupid questions.

Q1 What type of questions have you heard people ask about `BI SmartLINK`? What would you like to know more about `BI SmartLINK`'s behavior or functionality? What are things about the app that you would like to understand better? If nothing comes to mind now, no worries; we can come back to this towards the end of the interview.

Q2 What type of technologies do people describe being monitored by under ATD?

Q3 Based on your understanding, how do people:

(a) Begin to be monitored by the app?

(b) What reason are they given for why they have to use the app (vs being released etc)? Do they/you believe this reason to be true/valid?

(c) Have you heard of people getting their passports taken, as a threat of more monitoring?

(d) If you don't want to use the app, is the only alternative being detained?

(e) How long are people told they'll be monitored? How long does it actually end up being?

(f) Do they have to pay to use the app? Did they install it on their own phones or did they get a phone from someone else? Was that phone free? Was there any assistance offered with setup?

(g) Stop being monitored by the app? What led to this? What impact did this have on them?

Q4 What types of things do they do with the app? What type of changes (if any) have

you heard about how the app functions?

Q5 How did the app impact their behavior?

Q6 In general, what have you heard about people's experiences using the app? Think about a specific positive, neutral or negative experience you have heard and describe it to us.

Q7 Do you have any concerns about people using this app?

Q8 If you could give feedback to the app's developers, what would it be?

Q9 What have you heard about people's interactions with their case managers related to the app?

Q10 Have you heard of people experiencing technical problems with the app, such as faulty location detection?

Q11 Now that we have discussed the app thoroughly, do any additional questions come to mind that you did not previously mention?

Q12 What drew you to participate in this study?

Q13 What do you wish researchers would focus on or study?

Q14 We only have a few minutes left; is there anything else you want to share before the interview is over?

Demographics

I have a few demographic questions. Feel free to decline to answer them.

Q1 Please select your age range from the following: 18-24, 25-34, 35-44 ...

Q2 What is your gender?

Recruitment

Could you share the study with others who might be good participants? Here is a link to the screening survey and here is a link to the recruitment flier.

B.2 Codebook

Below is our codebook. We list our codes in the following format: `[theme:code]`.

- Power imbalances: Abuse by officials
- Power imbalances: Lack of accountability
- Power imbalances: Relationships with case specialists
- Negative impacts: Employment
- Negative impacts: Housing
- Negative impacts: Well-being/Stress from mistakes or bad advice
- Negative impacts: Passports confiscated
- Negative impacts: Increased surveillance
- Negative impacts: Usability problems impact compliance
- Shortcomings of ATD: Historic support for ATD
- Shortcomings of ATD: Not a real alternative
- Shortcomings of ATD: ATD has widened its net
- Shortcomings of ATD: ATD hurts compliance
- Shortcomings of ATD: Monitoring is becoming ubiquitous
- Recommendations: For developers
- Recommendations: For researchers
- Recommendations: For policymakers

- Questions about BI: Location tracking
- Questions about BI: Data practices

Appendix C

ADDITIONAL MATERIAL FOR DECEPTIVE DESIGN IN VOICE INTERFACES

C.1 Survey instrument

C.1.1 Screening questions

[Below we include the questions used to screen participants before continuing on to the main survey.]

There are many smart voice assistants. A smart voice assistant is something that responds to vocal commands or questions. For example, you could ask "what will the weather be tomorrow" or say "turn off the lights." Below are some examples of smart voice assistants. Smart voice assistants can run on a number of devices like smart speakers, appliances, TVs, etc. Which of these do you currently have on any of your devices? Select all that apply. [Response choices: ◯ Amazon Alexa ◯ Apple Siri ◯ Google Assistant ◯ Huawei Celia ◯ Microsoft Cortana ◯ Samsung Bixby ◯ None ◯ Other (free response)]

Which of these devices do you currently have at home? Select all that apply. [Response choices: ◯ Electric car (e.g., Tesla, Leaf) ◯ Smart bulb (e.g., Philips Hue, Wyze) ◯ Smart curtain (e.g., SwitchBot, American Homesupplier) ◯ Smart door/garage lock (e.g., August, Google Nest) ◯ Smart phone (e.g., Android, iPhone) ◯ Smart plug (e.g., Belkin/Wemo) ◯ Smart speaker (e.g., Google Home, Amazon Echo Dot) ◯ Smart tag (e.g., Apple Tags, Tile) ◯ Smart thermostat (e.g., Nest, Ecobee) ◯ Smart toy (e.g., Neurala, seebo) ◯ Smart TV (e.g., Roku, Apple TV) ◯ Smart watch (e.g., Fitbit, Apple Watch) ◯ Video camera / smart doorbell (e.g., Ring, Eufy Security)]

C.1.2 General usage questions

[At this point, if participants met our inclusion criteria (using a smart speaker) we invited them to participant in the larger, main portion of the survey. If they did not wish to participate they were directed back to Prolific and paid for the screening survey.]

What type(s) of smart speaker(s) do you currently use? Select all that apply. [Response choices: ◯ Echo Dot, Echo Plus, or Echo Flex ◯ Echo Show or Echo Spot ◯ Echo Look ◯ Amazon Tap ◯ Google Nest Audio or Mini or Google Home ◯ Google Nest Hub or Hub Max ◯ Apple HomePod Mini ◯ Sonos One or Move ◯ Bowers & Wilkins Zeppelin ◯ Other (free response)]

How many years have you had your smart speaker(s)? If you have multiple smart speakers, choose the longest duration. [Response choices: ◯ Less than 1 year ◯ 1-2 ◯ years ◯ 2-3 years ◯ 3-4 years ◯ 4-5 years ◯ 5+ years ◯ I'm not sure ◯ Other (free response)]

Which room(s) are your smart speakers stored in? Select all that apply. [Response choices: ◯ Living room ◯ Bedroom ◯ Bathroom ◯ Kitchen ◯ Office ◯ Family room ◯ Basement ◯ Dining room ◯ Main room (e.g., in a studio apartment) ◯ Other (free response)]

How frequently do you interact with your smart speaker(s)? [Response choices: ◯ Several times a day ◯ Once a day ◯ More than once a week, but not everyday ◯ Once a week ◯ Once a month or less frequently]

Please indicate your agreement with this statement: "I trust my smart speaker." [Response choices: ◯ Strongly agree ◯ Agree ◯ Neither agree nor disagree ◯ Disagree ◯ Strongly disagree]

In a few sentences, please explain why do you trust or distrust your smart speaker (free response).

What type of things do you use your smart speaker(s) for? Select all that apply. [Response choices: ◯ Playing music ◯ Controlling smart home appliances (e.g., lights, thermostat) ◯ Checking the weather ◯ Asking questions ◯ Setting timers/alarms ◯ Other (free response)]

On average, how many hours per day do you spend near your smart speaker in your home? By near, we mean close enough that you can activate it using your voice. [Response choices: 0, 1, ... , 17, 18+]

C.1.3 Scenarios

You will now be presented with three scenarios describing an interaction with a smart voice assistant on a smart speaker. As you read through each scenario, please read the text assigned to you aloud as if you are interacting with the smart voice assistant in real time. After listening to a short audio clip you will be asked a few questions about it. You must play the audio clip to advance to the next question.

[Participants were then randomly shown three of the 12 scenarios shown in Table 5.1. Below is an example with Scenario 1.]

First page:

Consider the following scenario in which you would like to cancel your subscription to a service:

You: "Voice Assistant, I'd like to cancel my premium subscription."

VA: [an embedded audio clip]

Second page:

(Attention check question) What were you trying to accomplish in the previous scenario? [Response choices: ◯ Starting a new subscription ◯ Canceling a subscription ◯ Getting a list of all subscriptions ◯ None of the above]

Third page:

You: "Voice Assistant, I'd like to cancel my premium subscription."
VA: [an embedded audio clip]

On a scale of very unproblematic to very problematic how would you rate this interaction? [Response choices: ◯ Very problematic ◯ Problematic ◯ Neither problematic or unproblematic ◯ Unproblematic ◯ Very unproblematic]

In a few sentences, please explain why you selected the above answer. (free response)

On a scale of very realistic to very unrealistic, how realistic do you think it is that a smart voice assistant might exhibit this behavior? [Response choices: ◯ Very realistic ◯ Realistic ◯ Neither realistic nor unrealistic ◯ Unrealistic ◯ Very unrealistic]

In a few sentences, please explain why you selected the above answer. Have you experienced something similar to this before? (free response)

C.1.4 Previous encounters with deception

Have you encountered any situations while interacting with your smart voice assistant, where you felt it was trying to trick, manipulate, or deceive you? For example, where you felt it was trying to trick, manipulate, or deceive you into granting a permission, sharing data, or making a purchase? (free response)

C.1.5 Participant behavior questions

Please answer the following questions honestly. Your answers will not affect your payment, approval status, or your future recruitment for our studies in any way.

Did you read the dialogue (i.e., anything that said "You:'..."') from the scenarios aloud as you went through them? [Response choices: ◯ Yes ◯ Sometimes ◯ No]

Did you repeat any of the dialogue from the scenarios to a smart speaker next to you to see what would happen? [Response choices: ◯ Yes ◯ Sometimes ◯ No]

Do you feel that you might have "cheated" anyway on this survey while taking it? If so, please add details below. This question is optional. (free response)

C.1.6 Demographic questions

Which language(s) do you use when you speak to your smart voice assistant (e.g., English, Spanish)? (free response)

How old are you? [Response choices: ◯ Under 18 ◯ 18-24 years old ◯ 25-34 years old ◯ 35-44 years old ◯ 45-54 years old ◯ 55-64 years old ◯ 65-74 years old ◯ 75-84 years old ◯ 85-94 years old ◯ 95+ years old ◯ Prefer not to say]

What is your gender? [Response choices: ◯ Woman ◯ Man ◯ Non-binary ◯ Prefer to self-describe (free response) ◯ Prefer not to say]

What is your race/ethnicity? Select all that apply. [Response choices: ◯ White ◯ Black or African American ◯ Middle Eastern or North African ◯ American Indian/Native American or Alaska Native ◯ Asian ◯ Native Hawaiian or Other Pacific Islander ◯ Hispanic, Latino, or Spanish Origin ◯ Other (free response) ◯ Prefer not to say]

What best describes your employment status? Select all that apply. [Response choices: ◯ Working full-time ◯ Working part-time ◯ Unemployed and looking for work ◯ Homemaker ◯ Stay-at-home parent ◯ Student ◯ Retired ◯ Prefer not to say]

What is the highest degree or level of school you have completed? [Response choices: ◯ No schooling completed ◯ Nursery school ◯ Grades 1-8—no diploma ◯ Grades 9-12—no diploma ◯ GED or alternative credential ◯ High school diploma ◯ Some college credit, but less than 1 year of college ◯ 1 or more years of college credit, no degree ◯ Associates degree (for example: AA, AS) ◯ Bachelor's degree (for example: BA. BS) ◯ Master's degree (for example: MA, MS, MEng, MEd, MSW, MBA) ◯ Professional degree beyond bachelor's degree (for example: MD, DDS,DVM, LLB, JD) ◯ Doctorate degree (for example: Ph.D., EdD) ◯ Prefer not to say]

How many people live in your household (including you)? [Response choices: 1, 2, ..., 9, 10+, Prefer not to say]

Which of the following best describes your educational background or job field? [Response choices: ◯ I have an education in, or work in the field of computer science, computer engineering, or IT ◯ I do not have an education in, or work in the field of computer science, computer engineering, or IT ◯ Prefer not to say]

Table C.1: Distribution of scenario responses. Participants (n=93) were randomly assigned three scenarios to evaluate.

Scenario	**# of Responses**
Scenario 1	23
Scenario 2	23
Scenario 3	24
Scenario 4	24
Scenario 5	22
Scenario 6	24
Scenario 7	20
Scenario 8	23
Scenario 9	25
Scenario 10	23
Scenario 11	24
Scenario 12	24

C.1.7 Feedback

If you have any feedback on this survey, please share it below. (free response)

C.1.8 Scenario responses

C.2 Participant demographics

Metric	Levels	Count
Gender	Woman	60
	Man	31
	Non-binary	1
	Questioning	1
Age	18-24 years	45
	25-34 years	93
	35-44 years	75
	45-54 years	30
	55-64 years	30
	65-74 years	6
Race/Ethnicity	White	76
	Asian	12
	Hispanic, Latino, or Spanish Origin	11
	Black or African American	3
	American Indian/Native American or Alaska Native	1
	Native Hawaiian or Other Pacific Islander	1
	Prefer not to say	1
Household size	One	15
	Two	24
	Three	23
	Four	21
	Five	7
	Six	2
	Seven	1
Employment status	Working full-time	50
	Student	14
	Unemployed and looking for work	12
	Working part-time	11
	Homemaker/Stay-at-home parent	10
	Retired	4
	Prefer not to say	2
Education	Grades 9-12—no diploma	1
	High school diploma	7
	Some college but no degree	18
	Associate's degree	11
	Bachelor's degree	34
	Professional/Master's degree	21
	Doctorate degree	1
Tech Background	No	76
	Yes	11
	Prefer not to say	6

Table C.2: Demographic information of the participants. Participants were able to select multiple levels for race/ethnicity and employment status.

Appendix D

ADDITIONAL MATERIAL FOR MODDED WHATSAPPS

D.1 Interview study participant demographics

Table D.1: Demographics of participants.

		No.	%
Gender	Female	10	50.0
	Male	10	50.0
Age	18 - 24	9	45.0
	25 - 34	11	55.0
Education	High school	8	40.0
	Diploma*	5	25.0
	Bachelors	5	25.0
	Postgraduate	2	10.0
IT/CS Background	Yes	8	40.0
	No	12	60.0
Employment	Student	7	35.0
	Employed	12	60.0
	Self-employed	1	5.0

*In Kenya, diplomas focus on practical skills and vocational training and are usually completed in 1 to 2 years.

D.2 Testing procedure

Below we outline our testing procedure for validating the mod features. We recorded the screens of the test phone and the normal phone using a third, recording phone placed on a camera stand.

Set-up

- Factory reset the test phone.
- Set the screen timeout to 30 minutes on all phones.
- Set-up the camera stand and the recording phone.
- Install Google Drive, Google Docs on the test phone.
- Install the target app onto the test phone: via ADB for modded apps (after enabling "USB debugging" in developer mode), and from Google Play for the official app.

Main procedure

- Start video recording.
- Open the target app and login.
- Enable settings related to the following features. If these settings are enabled by default or do not exist, make note of that.
  - Anti-Delete feature for messages
  - Anti-Delete feature for stories
  - Hide read receipts for messages
  - Hide read receipts for stories
  - Freeze "Last Seen"
  - Disable incoming calls

Message read receipt test

- Start the test.
- Test phone sends: 'Hi, test begins.'
- Normal phone reads the message;

- Test phone leaves chat and goes to main screen listing all chats
- Normal phone sends: 'Test received.';
- Test phone reads the message;
- Normal phone leaves chat and goes the main screen listing all chats
  - (Modded App Only) At this point, the normal phone should not have received a read receipt. If the result is different, note that.
- Test phone takes an image and sends it;
- Normal phone views the message and opens the image;
- Test phone leaves chat and goes to main screen listing all chats
- Normal phone takes an image and sends it;
- Test phone reads the message and opens the image;
  - (Modded App Only) At this point, the normal phone should not have received a read receipt. If the result is different, note that.

Status read receipt test

- Start the test;
- Test phone posts a status with an image;
- Normal phone checks the status;
  - Verify that normal phone's status view is visible on test phone.
- Test phone reads the message and opens the image;
  - (Modded App Only) At this point, the normal phone should not be able to see that the test phone viewed the status. If the result is different, note that.

Anti-deletion test (message, image, status)

- The conversation between the normal and test phone should be open on both devices
- Normal phone deletes a message "For Everyone";
  - (Modded App Only) We should still be able to see the message on the test phone. If not, note that.

- Normal phone deletes an image;
  - (Modded App Only) We should still be able to see the image on the test phone. If not, note that.
- Normal phone deletes status;
  - (Modded App Only) We should still be able to see the status on the test phone. If not, note that.
- Test phone deletes status;
- Test phone deletes message;
- Test phone deletes image;

Freeze "Last Seen" test

- Note down last seen time observed on test phone for normal phone:
- Close the app on the normal phone;
- Wait five minutes
- Note down last seen time observed on test phone for normal phone:
- Reopen app on normal phone and go to the chat with the test phone;
- Note down last seen time observed on test phone for normal phone:
- Note down last seen time observed on normal phone for test phone:
- Close the app on the test phone;
- Wait five minutes
- Note last seen time observed on the normal phone for test phone:
- Reopen app on test phone and go to chat with the normal phone;
- Note last seen time observed on normal phone for the test phone:

Disable Incoming Calls test

- Start test;
- Test phone calls the normal phone;
- Normal phone answers and hangs up;

- Normal phone calls the test phone;
- Test phone answers and hangs up;
  - (Modded App Only) The test phone should never receive this call. If it does then note this below.
- The procedure ends.

D.3 Network traffic tracker analysis

App ID / Domain name	WhatsApp	MOD1	MOD2	MOD3	MOD4b	MOD6	# of apps
google.com	X	X	X	X	X	X	6
whatsapp.com	X	X	X	X	X	X	6
whatsapp.net	X	X	X	X	X	X	6
android.com	X	X			X	X	4
applovin.com			X	X	X	X	4
flurry.com			**X**	**X**	**X**	**X**	**4**
googleusercontent.com			X	X	X	X	4
liftoff-creatives.io			X	X	X	X	4
liftoff.io			X	X	X	X	4
vungle.com			**X**	**X**	**X**	**X**	**4**
ad-tracker.network			X	X		X	3
adsmoloco.com				X	X	X	3
appcenter.ms			X	X		X	3
doubleclick.net			**X**	**X**		**X**	**3**
gbwhat.pro			X	X		X	3
gbwhatsapp.download			X	X		X	3
googleapis.com			X		X	X	3
gstatic.com			X	X		X	3
rethinkad.com			X	X		X	3
schemas.casa		X			X	X	3
sharenotes.co		X			X	X	3
sszqdpx.xyz			X	X		X	3
yandex.net			X	X		X	3
atomhike.com					X	X	2
google-analytics.com		**X**				**X**	**2**
googleadservices.com			**X**			**X**	**2**
googlesyndication.com			**X**			**X**	**2**
googletagmanager.com		**X**				**X**	**2**
i18n-pglstatp.com					X	X	2
inmobi.com					**X**	**X**	**2**
isnssdk.com					X	X	2
maticooads.com					X	X	2
mtgglobals.com					X	X	2
2mdn.net						**X**	**1**
adjust.com				**X**			**1**
alexmods.com		X					1
appsflyer.com					**X**		**1**
bearsplay.com					X		1
cmpc.fun						X	1
fouadmods.com		X					1
funsdata.com					X		1
googletagservices.com						**X**	**1**
justbigso.com					X		1
muyuekj0.com					X		1
pangle.io					X		1
pellturvy.com						X	1
rayjump.com					**X**		**1**
unity3d.com					X		1
vungle.io				X			1
yandexmetrica.com						X	1
ymetrica1.com						**X**	**1**
zmedia.vn						X	1
# of domains	4	10	21	21	28	40	
# of tracking domains	0	2	5	4	5	11	

Table D.2: Bolded domains names are the ones in the ad/tracker list [183] from Ublock Origin. The "X" mark indicates the network traffic for the app in a given column contained the domain name in the row.