Security and Privacy for Augmented Reality: Our 10-Year Retrospective

Franziska Roesner and Tadayoshi Kohno
Paul G. Allen School of Computer Science & Engineering
University of Washington
https://ar-sec.cs.washington.edu

1 Introduction

Almost exactly 10 years ago, in May 2011, we submitted our first paper on security and privacy for emerging augmented reality (AR) systems [15], to the 2011 USENIX Workshop on Hot Topics in Security (HotSec).[1,2] Though the paper was rejected — a revised version later appearing as the cover article of the *Communications of the ACM* magazine in April 2014 [16] — it launched us on a now 10-year research trajectory anticipating, studying, and designing to mitigate security, privacy, and safety issues in AR (and/or in MR, XR, VR[3]). Meanwhile, commercial AR/MR/XR/VR platforms have become available and continued to advance, including Google Glass in 2013, the Microsoft HoloLens and the Meta 2 in 2016, the Magic Leap One in 2018, the Microsoft Hololens 2 in 2019, and Facebook's Oculus Quest 2 in 2020.

In this paper, we reflect on our research agenda from 2011, summarize our work in this space since then, identify key remaining open problems, and make predictions for the next 10 years. Our goal here is to reflect on and summarize our own research trajectory, not to summarize or systematize the entire AR security and privacy research space. Consequently, we have not cited or discussed many excellent related works. We are very excited to see growing interest and work on AR/MR/XR/VR security, privacy, and safety (as exemplified by this very workshop), but we leave a broader systematization to a future effort.

[1] https://www.usenix.org/legacy/events/hotsec11/

[2] For the purposes of our retrospective and documenting this history, we have put our (rejected) May 2011 HotSec workshop submission online [15].

[3] We typically use the term AR for simplicity, but most of the issues we raise apply across the AR/MR/XR/VR spectrum.

1st International Workshop on Security for XR and XR for Security (VR4Sec), Co-located with SOUPS 2021, Virtual Event, August 2021.

2 Our Research Agenda and Results

In this section, we summarize our AR security research agenda and the progress we have made on it.

2.1 Identifying Risks and Challenges

A first key set of questions that our research has asked is around anticipating what could go wrong with emerging AR platforms and applications: What are the new security, privacy, and safety risks that will be created? What are the known risks that will be exacerbated? And what can we do about it?

Our 2014 CACM article [16] laid out an initial set of challenges and defined a research agenda towards security and privacy for AR systems. We characterized these challenges along two axes: systems scope (single applications, multiple applications within a single AR platform, and multiple communicating AR systems) and related to input, output, or data access. Not by coincidence, the rest of this paper is organized in a similar way.

More recently, in 2019 we convened academic and industry leaders in AR for an "Industry-Academia Summit on Mixed Reality Security, Privacy, and Safety" that we hosted at the University of Washington. As part of the resulting report [6], we revisited and refreshed key opportunities and concerns for emerging AR/MR/XR/VR systems. We also presented a threat modeling framework to help technology developers, researchers, and policymakers consider the potential risks and harms, as well as the potential benefits, with future designs.

Given these potential risks, our work has then looked towards system designs that can mitigate them. Specifically, we have been asking: how should AR platforms or operating systems be designed to mitigate security, privacy, and safety risks from potentially buggy or malicious applications [3]? That is, as shown in Figure 1, in our work we have typically considered applications (and sometimes users) to be untrusted, while considering the AR platform itself to be trusted and the focus of our design efforts.

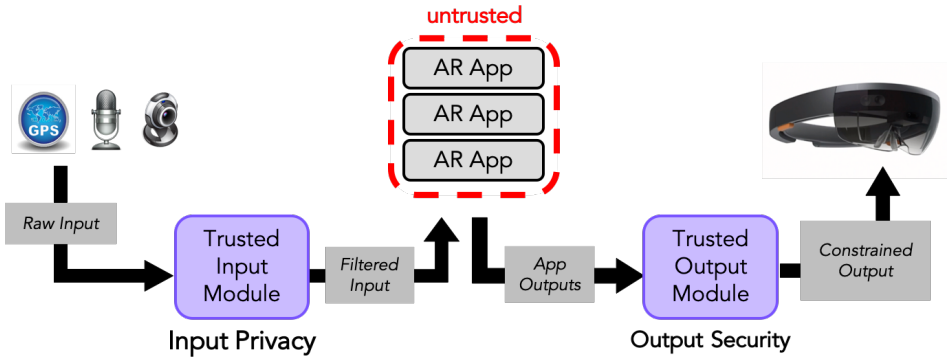

Figure 1: This figure shows how information flows through an AR platform as we have typically considered in our work: raw input is collected by sensors and passed to untrusted applications, which create virtual content that is output to users as audio, visual, or haptic feedback. Our (and others') work has introduced platform or OS level approaches to limit applications' access to raw sensor data (input privacy, Section 2.2) and to enforce policies on virtual content (output security, Section 2.3). AR platforms must also securely handle multiple apps running simulaneously (Section 2.4), as well as multiple users (i.e., multiple copies of this figure) interacting (Section 2.5). Non-users (i.e., bystanders outside of this figure) must also be considered (Section 2.6).

2.2 Input Privacy

Beginning with the input portion of the pipeline in Figure 1, we and other researchers considered how to mitigate privacy risks to AR users and bystanders from untrusted applications. Though such privacy risks are not fundamentally novel in AR settings, they require novel solutions and are significantly exacerbated by the need for always-on, continuously sensing technologies to support applications recognizing objects in and mapping virtual content to the physical world.

In an early collaboration with Microsoft Research, we proposed world-driven access control [17] as a novel access control model that limits AR applications' access to raw sensor data while relieving the user of the burden to make continuous fine-grained permission decisions. In world-driven access control, real-world objects can explicitly specify access policies, e.g., allowing the system to automatically stop recording in bathrooms or remove bystanders from video frames.

Beyond privacy, other input related challenges exist that we have not studied directly in the AR context, but that we anticipated in 2011 [15]: malicious real-world objects or content that aim to trick or exploit the AR system or its applications. Subsequent work on adversarial machine learning (including our own work on physical-world adversarial examples [5]) has shown that such attacks must be considered.

2.3 Output Security

Whereas input privacy is arguably a broader, existing concern (that also appears, for example, in smartphone and smart home contexts), the output side of the pipeline in Figure 1 presents fundamentally new and unique challenges in the AR context. Our research has sought to answer questions including: What risks are created by the virtual (audio, visual, and/or haptic) content displayed by untrusted applications via an AR device? How can AR platforms be designed to mitigate these risks?

Our work began by surfacing AR output risks posed by potentially buggy or malicious applications [10], which might (a) obscure another app's virtual content (to hide or modify its meaning), (b) obscure important real-world content (such as traffic signs or cars), or (c) disrupt the user physiologically (such as by startling them). We later explored (c) in more detail from a neuroscience perspective with collaborators from Meta [1], considering the potential threats from augmented reality to a user's perception, cognition, and motor responses.

To mitigate these risks, we proposed that AR platforms interpose on apps' requests to display virtual content (see Figure 1). To enable platforms to apply meaningful policies while still allowing apps the flexibility to display virtual objects embedded throughout the user's view of the physical world, we proposed three-dimensional "AR objects" as the granularity of display abstraction (rather than the two-dimensional windows or frames used in traditional computing platforms) [10]. We prototyped and evaluated these ideas in Arya, an AR platform that controls application output according to policies specified in a constrained yet expressive policy framework [12].

Multiple challenges remain towards realizing a flexible and secure trusted output module, including supporting more flexible policies, understanding and handling failures when policy enforcement relies on noisy sensor data input (e.g., recognizing physical world objects that trigger policies), handling non-visual virtual output (e.g., audio and haptic), and minimizing performance impacts. In addition, until recently we have only theorized the existence and impact of AR output security attacks; we are currently empirically (and ethically) studying their impact on real users.

2.4 Multiple Applications

Our work on input privacy and output security largely considered individual AR applications in isolation. However, we anticipate that future AR platforms will support multiple applications running — and modifying the user's perception of the physical world — simultaneously. We thus asked again: what security, privacy, and safety challenges will arise in multi-application contexts, and how should AR platforms be designed to support multiple (untrusted and mutually distrusting) applications?

We identified ways in which multiple applications might visually conflict with each other in malicious ways, and we explored the design space for how the AR platform or other stakeholders can manage conflicts between applications displaying content in a shared world [11].

Significant future work remains to be done on this research challenge. Our empirical investigation [11] and our other experiences with current commercial platforms suggest that multi-application support is still very limited. In our ongoing and future work, we are continuing to consider this challenge, e.g., building on ILLIXR [7].

2.5 Multiple Users

Just as multiple people share the same physical spaces without AR, we can expect (1) people to share the same physical space with AR and (2) people who are remote to share virtual content. Where we previously considered potentially malicious applications, in this context we must also consider potentially malicious *other users*. Again, we asked: what security, privacy, and safety challenges arise when multiple AR users interact, and how should multi-user support and sharing thus be designed?

In 2017, we conducted a user study in which we brought pairs of novice AR user participants into our lab to try the Microsoft HoloLens [13]. Among our findings, we learned that the fact that participants shared a physical space shaped their assumptions about shared virtual content; we saw participants engage in both conflicting and cooperating behaviors; we heard concerns about privacy and unwanted, harmful, and deceptive content; and we identified the need for access control for shared virtual content. These findings helped motivate our subsequent investigation into multi-user AR platform designs.

We then explored the requirements and design space for multi-user AR content sharing and access control [19]. We proposed the design of a multi-user AR sharing control module, which enables AR application developers to support users in controlling how they share AR content with others and how AR content is shared with them, while taking into account unique challenges in the AR context — most importantly, the integration of AR content with the physical world. We prototyped our design as ShareAR, a trusted application-level library for HoloLens [18].

ShareAR provides a set of composable sharing and access control primitives for use by applications, but because different applications may have vastly differently functionality needs and sharing semantics, it provides little guidance on what applications *should* do or how sharing and access control user interfaces and interactions should be designed. We are continuing to explore these questions with collaborators from the HCI community.

2.6 Bystanders

The people who use AR devices are not the only stakeholders who may be impacted: we must also consider the security and privacy implications for (and from) non-user bystanders.

Our work has surfaced the privacy concerns of bystanders [4,13] and identified potential design axes for privacy-mediating technologies [4]. Considering bystanders as a possible adversary, our work also demonstrated that AR headset displays can leak (potentially sensitive) visual information to people or sensors across the room [9].

Though the challenges are clear, and though our prior work has explored design axes [4] and particular solutions [17] for protecting bystander privacy, more work remains to be done towards solving this problem in practice.

2.7 Law and Policy

Alongside our technical research agenda, our work has sought to surface related policy and regulatory issues, and to inform policymakers about the considerations and risks with emerging AR technologies. For example, we identified hard problems of law and policy including around privacy, free speech, discrimination, and safety [14], which we further refined in a primer for policymakers [2]. Our recent Industry-Academia Summit Report [6] also includes policymakers among its intended audiences, though we did not have policy experts in attendance at the Summit and believe a deeper exploration from this perspective is still needed, particularly as the relevant technologies continue to advance and see adoption.

2.8 Envisioning

Throughout our 10-year research program, we have striven to envision possible futures and to explore AR security and privacy in such futures. Our 2011 HotSec submission included motivating fictional vignettes [15]. The *Our Reality* novella [8] surfaces issues relevant to security and privacy for AR, including questions about the role of online advertisements in AR environments, how digital content might be shared between users, privacy between users in an AR environment, and issues that can manifest when some but not all people have access to AR technologies.

3 Reflections and Challenges for the Future

Finally, we step back to reflect on our research trajectory over the last decade, and we identify key open challenges that we predict will become critical in the next decade.

Looking back, we have made substantial progress on the research agenda that we laid out starting in 2011. Yet many challenges remain to move these ideas from research to practice, as outlined in some of the earlier sections. Additionally, in some cases, our research has provided key insights and design explorations, but different AR platforms and contexts will require choosing different tradeoffs in practice. Finally, several wide open problems remain and/or have become visible to us over the years, as we discuss below.

Meanwhile, commercial AR/MR/XR/VR technologies have advanced significantly since our 2011 manuscript — indeed, the simple Google Glass heads-up display (released in 2013) had not even been publicly announced at the time — but non-smartphone form factors are still not widely deployed. With investments from major technology companies, we can expect continued advancements. However, current platforms have largely not yet addressed the security, privacy, and safety challenges that we identified in our work, still focusing primarily on hardware and use cases. We continue to argue that these security, privacy, and safety issues are existential concerns for these technologies, and that they must be considered well before widespread deployment, as solutions require answering fundamental platform design questions. In the last several years, we have been excited to see increased interest and investment in addressing these challenges among key industry players (e.g., those who attended our 2019 Summit [6]).

Looking ahead, we identify key areas where we expect AR advancements to raise critical security, privacy, and safety questions beyond those we have already considered. We anticipate developments and associated challenges due to **(a) market forces** (monetization and cross-platform support), **(b) increasingly immersive AR technologies** (physical world integration, as well as brain and body interfaces), and **(c) human factors in and around AR** (identity and disparate access).

1. **Monetization.** Companies that produce AR devices, and companies that create AR apps, may employ methods for monetization. These methods may involve tracking users and analyzing user behaviors, virtual (and user-targeted) ads embedded in the physical world, and attempts to modify people's behavior (e.g., guide them to specific stores). Exacerbating today's concerns with advertisements and tracking on the web, AR ads and tracking will raise substantial concerns about privacy, discriminatory targeting, and potentially problematic content integrated with a person's experience in the physical world.
2. **Cross-platform support.** If a single AR device does not (or should not) emerge as dominant in the market, then there will be a need to create and support cross-platform AR applications. We designed ShareAR [18, 19] with cross-platform support in mind. Security and privacy issues can arise if there are mismatches in expectations and assumptions at the interface between different systems.
3. **Physical world integration.** Companies will seek to further integrate AR devices with the physical world, such as by geolocating virtual content, or by collecting data to build rich maps of the physical world. AR devices may also integrate with nearby smart devices (such as Apple's AirTags), which may be designed specifically to facilitate new AR experiences. Designing these features in secure and private ways will raise challenges, e.g., how to manage the interaction of virtual content and physical-world data collection with the ownership of physical space (discussed more in our Summit report [6]).
4. **Interfacing with the brain and body.** Future AR technologies may explicitly interface with the body and brain, with sophisticated body-sensing and brain-machine interface technologies. Further, the immersive nature of AR may create new opportunities for adversarial applications to influence a person's thoughts, memories, and even physiology. While we have begun to explore the relationship between AR technologies, neuroscience, security, and privacy [1], much more work needs to be done to both understand the risks and to mitigate them.
5. **Social interactions and identity.** In social AR applications, a tension between identity and anonymity may emerge. Will users of an AR application be able to know, with confidence, the identity of the person they are interacting with, or will it be possible to assume someone else's identity (akin to phishing attacks or fake social media accounts on the web)? Alternately, if desired, will AR users be able to be anonymous? How can anonymity and accountability for undesirable behaviors be balanced?
6. **Disparate access.** Not everyone may have the same access to AR technologies. Some people may not be able to afford it; other people may not have access due to a technology's poor accessibility capabilities. Building on our prior work on security and privacy concerns for bystanders [4], the design of future AR technologies must consider the security and privacy needs of non-users.

We are excited to see the research community around security, privacy, and safety for AR/MR/XR/VR growing, and we look forward to seeing and contributing to the next 10 years of work in this space.

Acknowledgements

We are grateful to our many collaborators over the past decade, listed as co-authors in the references below — particularly David Molnar, then at Microsoft Research, who helped us kick off this research vision, and the current and past students at the University of Washington who have contributed or are contributing significantly to this research agenda (and

gave feedback on a draft of this retrospective): Kaiming Cheng, Kiron Lebeck, and Kimberly Ruth. This work has been supported in part by the National Science Foundation under Awards CNS-0846065, CNS-0905384, CNS-1513584, CNS-1565252, CNS-1651230, an NSF Graduate Research Fellowship under Grant DGE-0718124, Microsoft Research and a Microsoft Research Fellowship, a Washington Research Foundation Fellowship, the Short-Dooley Professorship, the UW Reality Lab, the UW Tech Policy Lab, an "Explorations of Trust in AR, VR, and Smart Devices" grant from Facebook, and a gift from Google.

References

[1] Stefano Baldassi, Tadayoshi Kohno, Franziska Roesner, and Moqian Tian. Challenges and New Directions in Augmented Reality, Computer Security, and Neuroscience – Part 1: Risks to Sensation and Perception. Technical Report arXiv:1806.10557, 2018.

[2] Ryan Calo, Tamara Denning, Batya Friedman, Tadayoshi Kohno, Lassana Magassa, Emily McReynolds, Bryce Newell, Franziska Roesner, and Jesse Woo. Augmented Reality: A Technology and Policy Primer. Technical report, Tech Policy Lab, University of Washington, 2015.

[3] Loris D'Antoni, Alan Dunn, Suman Jana, Tadayoshi Kohno, Benjamin Livshits, David Molnar, Alexander Moshchuk, Eyal Ofek, Franziska Roesner, Scott Saponas, Margus Veanes, and Helen J. Wang. Operating System Support for Augmented Reality Applications. *Workshop on Hot Topics in Operating Systems (HotOS)*, 2013.

[4] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. In Situ with Bystanders of Augmented Reality Glasses: Perspectives on Recording and Privacy-Mediating Technologies. In *ACM CHI Conference on Human Factors in Computing Systems*, 2014.

[5] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. Robust Physical-World Attacks on Deep Learning Visual Classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[6] Franziska Roesner and Tadayoshi Kohno (editors). 2019 Industry-Academia Summit on Mixed Reality Security, Privacy, and Safety: Summit Report. Technical report, University of Washington, 2020.

[7] Muhammad Huzaifa, Rishi Desai, Samuel Grayson, Xutao Jiang, Ying Jing, Jae Lee, Fang Lu, Yihan Pang, Joseph Ravichandran, Finn Sinclair, Boyuan Tian, Hengzhi Yuan, Jeffrey Zhang, and Sarita V. Adve. Exploring Extended Reality with ILLIXR: A new Playground for Architecture Research. Technical Report arXiv:2004.04643, 2021.

[8] Tadayoshi Kohno. *Our Reality: A Novella*. 2021.

[9] Tadayoshi Kohno, Joel Kollin, David Molnar, and Franziska Roesner. Display Leakage and Transparent Wearable Displays: Investigation of Risk, Root Causes, and Defenses. Technical Report MSR-TR-2015-18, Microsoft Research, 2015.

[10] Kiron Lebeck, Tadayoshi Kohno, and Franziska Roesner. How to Safely Augment Reality: Challenges and Directions. In *Workshop on Mobile Computing Systems and Applications (HotMobile)*, 2016.

[11] Kiron Lebeck, Tadayoshi Kohno, and Franziska Roesner. Enabling Multiple Applications to Simultaneously Augment Reality: Challenges and Directions. In *Workshop on Mobile Computing Systems and Applications (HotMobile)*, 2019.

[12] Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. Securing Augmented Reality Output. In *IEEE Symposium on Security & Privacy*, 2017.

[13] Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. Towards Security and Privacy for Multi-User Augmented Reality: Foundations with End Users. In *IEEE Symposium on Security & Privacy*, 2018.

[14] Franziska Roesner, Tamara Denning, Bryce Clayton Newell, Tadayoshi Kohno, and Ryan Calo. Augmented Reality: Hard Problems of Law and Policy. In *UbiComp Workshop on Usable Privacy & Security for wearable and domestic ubIquitous DEvices (UPSIDE)*, 2014.

[15] Franziska Roesner, Tadayoshi Kohno, and David Molnar. Augmented Reality: Challenges and Opportunities for Security and Privacy, May 2011. Manuscript, https://ar-sec.cs.washington.edu/files/ARSec2011.pdf.

[16] Franziska Roesner, Tadayoshi Kohno, and David Molnar. Security and Privacy for Augmented Reality Systems. *Communications of the ACM*, 57(4):88–96, 2014.

[17] Franziska Roesner, David Molnar, Alexander Moshchuk, Tadayoshi Kohno, and Helen J Wang. World-Driven Access Control for Continuous Sensing. In *ACM Conference on Computer and Communications Security (CCS)*, 2014.

[18] Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. ShareAR: Secure and Private AR Sharing Toolkit. https://arsharingtoolkit.com/.

[19] Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. Secure Multi-User Content Sharing for Augmented Reality Applications. In *USENIX Security Symposium*, 2019.