2019 INDUSTRY-ACADEMIA SUMMIT on

# MIXED REALITY

## Security, Privacy, and Safety

SUMMIT REPORT


SECURITY & PRIVACY RESEARCH LAB
UNIVERSITY of WASHINGTON

UW REALITY LAB

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

# TABLE OF CONTENTS

# 1 PREAMBLE

**Augmented and Mixed Reality technologies bring great potential benefits, but they also raise new and serious computer security, privacy, and safety risks. Because AR/MR technologies have a unique ability to directly and immersively impact a user's perceptions of and interactions with the physical world, these risks can be fundamentally different from the risks with other technologies. Moreover, these risks—and consumers' potential concerns about these issues—may pose an existential threat to the widespread adoption and success of these technologies.**

Co-funded by the UW Reality Lab and the UW Security and Privacy Research Lab in the Paul G. Allen School of Computer Science & Engineering, in Fall 2019 the University of Washington hosted a Summit that brought together field leaders across industry and academia, to discuss and formulate visions on how to achieve the full benefits of AR and MR while mitigating security, privacy, and safety risks. A PDF of this report, and additional information about the UW efforts on AR/MR + security, can be found at https://ar-sec.cs.washington.edu/.

This report serves as a written record of key, publicly-discussable aspects of the Summit. Our goal was to bring together representatives from different companies and different perspectives for a joint conversation in a neutral academic context. To facilitate open discussion, the Summit was held under the Chatham House Rule, and so participants and their organizations are named in this report only with explicit permission. Participants represented a variety of academic and industry backgrounds, with many years of experience and expertise in AR/MR, computer security and/or privacy, and related areas. Participants who chose to be named are listed in Section 7. The ideas in this report are curated from discussions among all the participants, including those not listed in Section 7; at the same time, the ideas and opinions captured here do not necessarily reflect the opinions of all participants (nor their organizations).

**The Summit began with several invited keynote presentations:**

- Blair MacIntyre (Mozilla and Georgia Tech) and Diane Hosfelt (Mozilla) gave a presentation titled "The Immersive Web: Private, Safe, Secure and Ethical";
- Justin Quimby from Google gave a presentation titled "AR in Google Maps";
- Kimberly Ruth from the University of Washington gave a presentation titled "Understanding and Designing for Security and Privacy in Multi-user AR Interactions";
- Kate McKinley from Facebook gave a presentation titled "What Keeps Me Up at Night".

**Following these context-setting talks, participants worked in small groups to:**

1 Identify key opportunities with AR/MR, as well as key challenges to addressing computer security, privacy, and safety in AR/MR systems;

2 Explore specific methods for further defining and/or mitigating those challenges.

For (1), we summarize the key topics of discussion below, in Section 2. For (2), based on participant interest, one sub-group explored a general framework for AR/MR designers to consider potential risks (Section 3), and another sub-group conducted a deep dive into one specific class of challenges related to physical location (Section 4).

2

# KEY OPPORTUNITIES AND CONCERNS

**We begin by summarizing key conversation points that emerged during the initial small-group discussions — about what could "go right" or "go wrong" with AR/MR technologies, as well as other issues, tradeoffs, and design considerations that must be considered. This section highlights points that at least one participant mentioned, but does not necessarily reflect the beliefs or consensus of all participants.**

**TABLE 1: Summary of key opportunities and concerns.**

| What could go right | What could go wrong |
|---|---|
| Desirable AR/MR functionality and apps<br>(e.g., "contextual magic, "super powers", positive multi-user interactions)<br>Getting security, privacy, safety right | Considering security, privacy, safety too late<br>Physical harm<br>Undesirable virtual content<br>Manipulating reality<br>Privacy violations<br>Multi-user challenges<br>Too much or too little advertising<br>Technology dependency<br>Failing to consider all types of stakeholders<br>Unexpected broader societal impact<br>Traditional computer security issues<br>Safeguards preventing desirable functionality |

## IMPORTANCE OF ACTING NOW

One key concept that emerged during the first breakout session was the importance of acting now, to minimize the likelihood of any "go wrong" situation emerging. By acting thoughtfully and proactively now, there is the potential to "get it right" from the start. As one motivation, participants observed that the Web grew organically over decades, with a (largely) "see what sticks" approach, rather than a principled, security-driven approach from the beginning. The result, for the Web, was many years of pervasive computer security concerns, and a Web today that is complicated to reason about from a security and privacy perspective. Though it is of course hard or impossible to evaluate hypotheticals in hindsight, some participants found this metaphor motivating to consider security and privacy issues in AR/MR *before* the technologies are widely deployed and designs are hard to change due to a need for backwards-compatibility.

Indeed, some participants wondered if there will be a catalyzing event — such as a major compromise, or significant harm to some users — that would result in significant broader industry movement in this space, or whether there would be gradual, slow but continuous progress over time. Related to a catalyzing event, participants suggested that companies should consider how to prepare for and respond to such an event, prior to it happening. Further, participants suggested that it would be desirable to make significant progress on proactive security, privacy, and safety mechanisms before any catastrophic catalyzing event occurs.

## OPPORTUNITIES: WHAT COULD GO RIGHT

**Desirable AR/MR functionality and applications.** Some discussions focused on the many positive opportunities afforded by AR/MR, independent of security, privacy, and safety. AR/MR has the potential to bring significant value to users and to society, and the attendees knew this. For example, attendees talked about AR/MR technologies that implement "contextual magic" and that provide immersive content with seamless user engagement. Participants observed that AR/MR technologies could help users overcome human limitations of time and space, giving users "super powers" (but explicitly not in "creepy" ways), and in ways that are accessible to everyone, regardless of physical or financial or other capabilities. Participants

discussed numerous applications, ranging from navigation and maps to social applications. Participants envisioned a future with rich multi-user AR/MR functionality as well as interoperability between AR/MR platforms.

**Getting security, privacy, safety right.** Other discussions focused specifically on what could "go right" with respect to security, privacy, and safety. One participant observed that an ideal world might be one in which users and developers have the control that they want, and feel the agency to use those controls. Other positive, possible futures include ones in which the data collected by AR/MR systems and applications is only used for what users want and expect, where users can know for sure that this property holds true, where AR/MR systems and applications are held accountable for data use, and where users can view and edit data about themselves. Participants suggested that a "go right" future might include one in which the well-known computer security principle of "least privilege" is deployed ubiquitously, thereby ensuring that systems do not have access to data that they do not need, and that only data that is later needed is stored. Participants also suggested that a positive future could include one in which bystanders (those near another's AR/MR device) can control what someone else's AR/MR device records about them. A positive future might also provide perfect authentication for virtual photorealistic actors, thereby allowing an AR/MR device to help users distinguish between real actors and adversarially-created "deep fakes". A positive future might also include sensible and usable government policies.

## CHALLENGES: WHAT COULD GO WRONG

Given the focus of the workshop, most discussions surfaced numerous examples of how things could "go wrong". This is not to say that participants thought that all these situations *would* necessarily arise in the future, but that they *could* potentially arise if the technology continues to progress without any mitigations, safeguards, or proactive planning.

**Physical or psychological harm.** One example class of "go wrong" situations included physical harm (e.g., to people, animals, or the environment), or psychological harm to people. For example, a buggy or malicious AR/MR application could startle a user or display virtual content that obscures their view of crucial physical-world objects (e.g., an oncoming car). These risks have been discussed in prior academic research (e.g., Lebeck et al. 2017, Baldassi et al. 2018).

**Undesirable virtual content**. Participants also considered the risk of the creation of AR/MR content that users do not want to see. The impacts could range from harassment (e.g., problematic AR/MR graffiti on real-world objects or in public spaces, or offensive AR/MR content displayed near individuals), to AR/MR-based phishing schemes, to undesired AR/MR advertisements and spam.

**Manipulating reality.** Another class of "go wrong" situations involve the use of AR/MR technologies to interfere with people's perceptions of reality. For example, a subtle manipulation might result in a $20 bill looking like a $5 bill. More extensive manipulations could result could lead to different people seeing different realities — resulting in "filter bubbles" (in which people only see information that reinforces their existing worldview), blindness to certain aspects of reality (e.g., an AR/MR technology could remove real-world items that the user does not want to see, or that the app designer does not want them to see), or the spread of immersive disinformation.

**Privacy violations.** Privacy is often raised as a potential concern with AR/MR technologies, since these systems fundamentally need significant access to sensor data about the physical world and the user to fulfill their intended functionalities. This topic came out during our discussions as well. For example, AR/MR platforms might collect continuous fine-grained and/or biometric data about a user, including eye tracking and heart rate data, that could be used to make sensitive inferences about a person — not to mention the collection of continuous video and audio data about the user and bystanders.

Moreover, participants asked whether the permanence of data could pose long-term problems for users. Many of today's computer systems have the potential to store significant amounts of data about a person, over the person's lifetime. However, those computer systems have traditionally been ones that only glimpse a portion of the user's life. AR/MR technologies have the potential to be much more personal — worn and used by a person as they go about all aspects of their lives — and hence have the potential to store significantly more information about a person.

**Multi-user challenges for virtual content.** There was a rich discussion of the visibility of virtual AR/MR objects, the question of who can control or edit those objects, the ownership of those objects, and the ownership of physical spaces. Regarding the sharing of virtual AR/MR objects (to allow, for example, other people to see or edit the objects), participants observed that, compared to sharing simple content like photos or documents, AR/MR content may

be significantly more complex, and hence significantly harder for users to reason about. Participants observed that without clear ownership rules for physical spaces, adversarial users could (for example) post virtual graffiti around the world, even in places where cultural norms or rules would not allow them to place or manipulate physical-world objects. Participants dug deeper into these issues in a later breakout, which we summarize in Section 4.

**Too much or too little advertising.** Participants also discussed the role of advertising in future AR/MR systems. One question was whether advertising would support the future AR/MR ecosystem, much like advertising helps support many Web entities today. Exploring a sequence of hypotheticals, participants discussed whether giving users full control over the display of AR/MR content in their environments could lead to the inability for advertisers to display AR/MR advertisements, which could then lead to hurting the AR/MR advertising industry, which could then in turn significantly harm the AR/MR industry as a whole. Participants wondered whether advertisements could be limited to certain contexts (e.g., only while a person is doing certain leisure activities). Participants also wondered how much private information about a person's activities or environment would flow to the advertisers, to enable the advertisers to create extremely well-targeted personalized ads.

**Technology dependency.** Another class of "go wrong" situations revolved around technology dependency. For example, participants considered a situation in which people became reliant on AR/MR technologies, and then the technology stopped working, perhaps because of a denial-of-service-style attack. Participants also asked whether security, privacy, or safety issues could be amplified if a user was not able to take off or turn off the device (e.g., a car windshield).

**Failing to consider all types of stakeholders.** Participants also warned that AR/MR technologies might not sufficiently consider the full set of stakeholders that could be impacted. Participants observed that different populations might have different needs, or might need to be considered differently. One example population is children, who might not have full autonomy or might have different legal protections. Another population is medical patients, who might be using an AR/MR device for healthcare purposes, thus potentially putting the AR/MR device under HIPAA regulations (in the U.S.). Another population might be people with disabilities, who might benefit from novel accessibility-focused interfaces.

It was also pointed out that different countries and cultures have different norms and expectations regarding personal space and privacy. For example, consider two users interacting via an AR/MR system who have significantly different cultural norms or expectations regarding physical space (e.g., with different expectations about who can place virtual objects where). As another example, consider two people interacting in a physical environment where only one of the two can afford an AR/MR device: the bystander would be excluded from virtual content that the AR/MR device owner and others can create and see, while also being subject to extensive data collection by the AR/MR device.

Participants further discussed the challenges with trying to simultaneously support such a diverse collection of stakeholders. For example, drawing from past work (Jang et al. 2014), it was pointed out that accessibility interfaces may expose AR/MR systems to certain threats, thereby making it challenging to provide both certain security, privacy, and safety properties and certain accessibility properties at the same time. As another example, participants discussed the value of having strong user identification, perhaps even requiring all users to prove their identity and provide a residential address for verification. However, such a requirement could impact the use of AR/MR technologies by activists, who might desire anonymity, or by homeless people or others without a permanent residential address.

**Unexpected broader societal impact.** Participants discussed the potential for AR/MR technologies to impact society, and the importance of considering such impacts early. For example, suppose a navigation application incorporated safety information. All users of the AR/MR navigation technology might thus avoid certain areas of a city, thereby potentially impacting both the areas of the city that are traversed and the areas of the city that are avoided. As another example, consider how different cultures have different notions of personal space. If an AR/MR technology is designed around one of those cultural norms, participants observed that the use of the technology by people of other cultures could eventually result in cultural changes.

**Traditional computer security issues.** While a majority of the conversation focused on the novel interface between the user and the AR/MR systems, participants stressed the importance of considering the traditional computer security of the supporting infrastructure as well. For example, a compromise of a supporting server could lead to significant data exposure, or to the ability for a hacker to adversarially manipulate the state of an AR/MR system.

**Safeguards preventing desirable functionality.** Finally, participants also considered situations in which security solutions or regulations might prevent certain desirable functionality or innovation. For example, there may be valid reasons for users to have multiple personas in an AR/MR system, but strong biometric capabilities, which could assist with strong user authentication and enhance security from certain perspectives, could also have the negative consequence of preventing users from having multiple personas. As another example, if controls are built into a system to prevent AR/MR systems from recording video content in certain environments or in certain situations, then those restrictions could have the negative consequence of preventing recording in situations that society would believe should be exceptions (e.g., recording of abuse of power by governments or recording an ongoing crime).

Fundamentally, participants observed that if AR/MR systems try to prevent "bad" situations from arising, they may also inadvertently limit "good" uses of the AR/MR technologies. On the other hand, if AR/MR systems do prioritize all the "good" uses of AR/MR technologies, then they may not sufficiently protect against such "bad" situations. Thus, both too much and too little attention to security, privacy, and safety could lead to "go wrong" situations.

## TOWARDS SOLUTIONS

We highlight several discussion points that arose when participants thought about how to move towards ways to address the above challenges.

**Trust and responsibility.** Participants observed that there are many entities whom users and bystanders will need to trust, in order to have confidence that an AR/MR system provides certain security, privacy, or safety properties — and that this trust must be established early for the technology to see widespread success. Possible entities to trust include the hardware provider, the AR/MR operating system provider, the application providers, other users, or a collection of some or all of these entities, or other entities. A key question emerged regarding how users can determine that these entities are actually trustworthy, and that any promised protection mechanisms (technical, regulatory, or otherwise) are actually in place and achieve the intended security, privacy, or safety goals.

Another way to consider who users must trust is to ask who should be responsible for addressing security, privacy, and safety in future AR systems. Possible responsible parties included the AR/MR device manufacturer, the developer of applications, legislators, a collection of some or all of these parties, or others.

**Industry standards.** For example, participants discussed the role of industry standards, and whether industry standards could help ensure progress across all AR/MR systems. Participants suggested that while it is important to focus on this issue now, it might also be too early for standardization. Participants suggested that there might need to be significant experimentation and failures first, before informed standards can emerge. Participants asked whether there were low-hanging research efforts, perhaps manageable in the one-year timeframe, that could help significantly advance the field's understanding at this time. For example, some participants observed that it might be valuable to explicitly identify different classes or levels of AR/MR technologies, and then focus different solutions at these different levels. This conversation drew inspiration from SAE International's five different levels of automation

for autonomous vehicles, which inform guidance and policies for vehicle manufacturers.[1] In the AR/MR space, one class might be that of single-user, single-platform AR/MR technologies with a single application running at a time. Another level might be that of multi-user, cross-platform AR technologies with multiple simultaneously-running applications.

**Regulation and policy**. Participants also discussed the potential role for regulation. It was observed that even if some companies implement strong security, privacy, and safety protections, regulation could have a role in helping make sure that all companies do so. It was also observed that without regulation, users may not trust or believe that the companies implement these protections. However, participants also urged caution, noting the premature legislation might not reflect the full richness and complexity of the AR/MR space. For example, it might seem natural for legislation to say that if one owns a physical space, then they should own the digital content in that space as well. Such legislation, though, would create significant technical challenges given the complexities with space ownership verification. At the same time, flimsy regulations full of loopholes could lure users into a false sense of security. Thus, while legislation could prove valuable, participants observed that it is essential to be thoughtful about the legislation, not rush into legislation early, and consult domain experts during the legislation process.

**Usable controls for users.** Participants observed that giving users controls for various settings or actions is different than giving participants usable controls. Numerous ideas were surfaced related to usability. For example, one participant observed the success of "airplane mode" for phones, and suggested an airplane mode for AR/MR device privacy. Participants observed the importance of giving users the ability to "return to reality" if, for example, the AR/MR system is misbehaving or if adversarial content appears. Participants also observed the benefit of "user-driven access control" (Roesner et al. 2012), whereby the AR/MR system is capable of interpreting and acting on the user's natural gestures; to be successful, the gesture must be clear and unambiguous, however. Participants observed that application permission management is notoriously difficult, even for more mature technologies like smartphones or browser extensions; managing permissions for AR/MR applications should build on the knowledge and experiences from these other domains. Participants also asked whether an AR/MR system's design could leverage real-world metaphors and, in doing so, enable more intuitive controls (e.g., lowering a window shade might be a physical-world analogy for desiring more privacy). Related prior work has considered gestures allowing users to indicate private areas of the physical world (Raval et al. 2016).

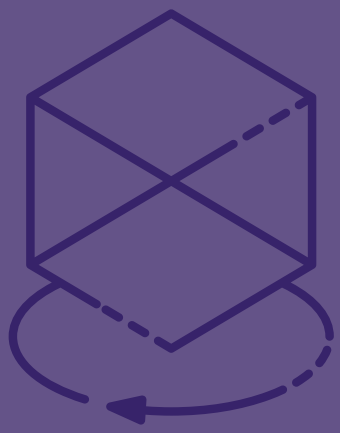
**Support for application developers.** Participants also highlighted the importance of considering usability for AR/MR application developers, enabling them to support specific security, privacy, and safety goals. For example, a security- and privacy-focused toolkit, which considers both developer and end-user usability, could help provide uniformity across apps, and could help developers create applications with certain security properties, even if the application developers are not security experts themselves. For example, it was observed that built-in security in Web frameworks can help non-security experts build websites that are more resilient to certain classes of attacks.

**A "Bill of Rights for Digital Spaces".** Finally, participants asked whether there is the potential for a "Bill of Rights for Digital Spaces", which might outline what a person might reasonably expect when interacting in an AR/MR digital environment.

[1] https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles

3

# A FRAMEWORK FOR CONSCIOUS AR/MR DESIGN



**A classic and crucial technique in computer security is threat modeling: systematically considering assets that should be protected in a system, and how (and in the face of which adversaries) that system might be vulnerable. In this section, we propose one specific approach for applying this type of thinking to mixed reality design. Specifically, a subset of the Summit attendees developed a potential "Fill in the Blanks" framework for designers to use when creating new AR/MR technologies, to scaffold careful consideration of the potential risks and harms, as well as the potential benefits, with these designs. The goal of this framework is to provide a way to help designers, engineers, policymakers, and researchers communicate with each other about mixed reality platforms and applications — and their associated potential benefits and harms — in a common language.**

**SPECIFICALLY, WE ENCOURAGE DESIGNERS TO THINK ABOUT BOTH *CONTEXT* AND *INTERACTIONS:***

**A key question for considering potential *context:***
"What happens when **[entities]** can **[process/action]** using **[sensors/input]** in **[setting]**?
The outcome could then have the following **[benefit]** and/or **[harms]**."

**A key question for considering potential *interactions*:**
"What happens when **[entities]** **[action]** **[entities]** using **[sensors/input]** in **[setting]**?
The outcome could then have the following **[benefit]** and/or **[harms]**."

These questions highlight the importance of understanding and accounting for the *entities*, *processes and actions*, *sensors*, and *context*. The answers to the questions will then reveal the *benefits* and *harms*. Below, we provide examples (but not an exhaustive list) to consider for each of these categories, as well as examples of how to use the fill-in-the-blank framework to surface potential benefits and harms.

**Entities,** or stakeholders, are any parties who may be directly or indirectly involved in or impacted by the design, implementation, and/or use of a technology. Examples include: the user themselves (including diverse types of possible users), the device owner, the platform or operating system designer(s), app developers, bystanders, content providers, advertisers, enterprises or corporations, stalker or harassers, friends or family members or colleagues of the user, other AR/MR users, law enforcement, governments, nation state actors, communications service providers, and the owners of physical-world spaces.

**Processes and actions** refers to what an entity might do — for example, how and where it might collect, store, or use data, and/or what and how it might produce content for the user. Possible actions might include but are not limited to: detecting spatial position of users and objects, recognizing physical objects in the world, predicting future body movements of the user, recognizing/predicting/inducing the user's body positions, creating a three-dimensional reconstruction of a space, displaying virtual content overlaid on the user's view of the physical world, augmenting a real person's facial or body features, and inferring information about other applications the user is running.

**Sensors and other inputs** to an AR/MR device or application might include but are not limited to: inward and outward facing cameras, RGB cameras, IR cameras, microphones, eye tracking, depth sensing, gyroscopes, accelerometers, temperature sensors, GPS, cellular, electrodermal, Wifi, Bluetooth, heart rate, brain activity, and pressure and touch sensors.

**Setting** refers to the way or context in which an AR/MR device or application is used. This can refer to both the physical instantiation of the technology (e.g., wearable headset vs. on a phone vs. in an automotive windshield) as well as where the technology is used (e.g., at home, in the car, in a workplace, in a public space, in a military setting). Given the setting, there may be different constraints on system use, different potential risks, different potential benefits, and different potential solutions. For example, a wearable can be more easily removed than a system in a car windshield.

**Benefits and harms** may be further explored by considering the following questions:

- Which issues are unique and/or new with AR/MR technologies? And which issues are significantly exacerbated in an AR/MR context, even if they have preexisting analogues?
- Who is affected: the AR/MR user, another AR/MR user, non-user bystanders, the person or entity who owns a physical object or space?
- What is the degree and type of harm? For example, physical harm, psychological harm, or financial harm.
- How permanent is the harm? Is the harm something that happens temporarily, or does it lead to a long-term impact on someone?

## EXAMPLES

Example explorations that the proposed framework might surface include:

- What happens when a third party app can infer a person's preferences using eye-tracking data in a shopping mall? A benefit to the AR/MR user could be personalized recommendations. Harms to the user could include the risk of biased advertising and/or pricing. While these issues exist on the Web today, they could be greatly exacerbated by using rich sensor data and placing advertisements/prices into the user's physical world.
- What happens when a company can infer visuomotor health markers using eye-trackers and inertial sensors (accelerometers/gyroscopes)? A benefit to the AR/MR user could be early diagnosis of neuromotor degenerative disorders. Harms to the user could be social/emotional anxiety, or undesirable impacts on employment or insurance. As in the previous bullet, these benefits and harms are significantly exacerbated due to the rich data collection in AR/MR or related technologies, compared to, say, smartphones.
- What happens when AI can model a user in AR/MR or virtual reality (VR) using accelerometers, gyroscope, and rich video data in a collaboration/meeting setting? A benefit to all users could be to "smooth out" the avatar's movements during low network reliability. A harm to any user could be that an attacker can replicate a user in VR (i.e., impersonation or identity theft, "deep fakes"). While such attacks have pre-AR/MR analogues, they may be significantly more effective in immersive environments.
- What happens when spatial software can redirect a user's movement through space? A benefit to the AR/MR user could be the ability to create an infinite virtual space for users to naturally walk and explore in a limited physical space. A harm to the user could be that malicious, malfunctioning, or compromised software can induce vestibular mismatch and/or direct the user to a dangerous location. These issues are unique or new to AR/MR technologies.
- What happens when users/companies can place 3D content in real-world locations that are not owned by the user? A benefit could be that those who don't own property have the ability to make changes to the world. A harm could be that property owners can have content placed in their locations that they don't approve of (e.g., Pokémon Go players going into the Holocaust Museum to catch Pokémon[2]). These issues — many of which are unique to AR/MR — form the motivation for our deep dive on complexities with ph ysical space in the next section of this report.

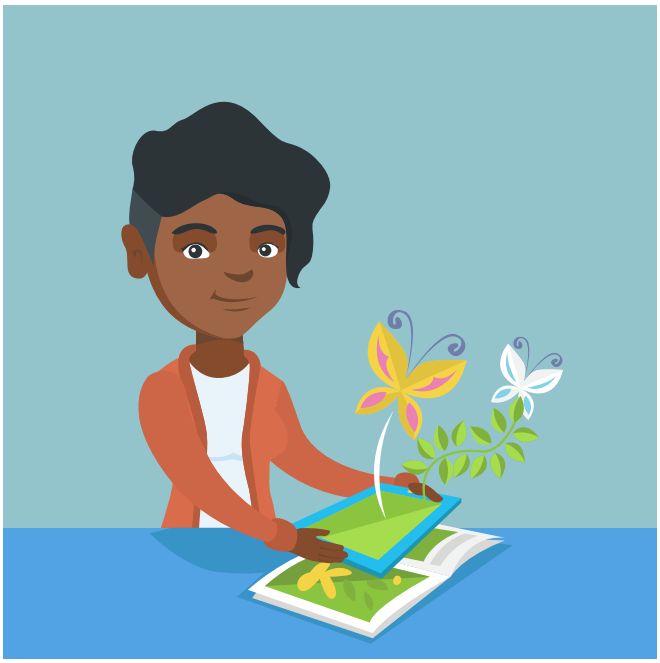
[2]https://www.washingtonpost.com/news/the-switch/wp/2016/07/12/holocaust-museum-to-visitors- please-stop-catching-pokemon-here/

4

# DEEP DIVE: MANAGING VIRTUAL CONTENT AND PHYSICAL SPACES

**Over the course of the Summit's breakouts, about half of the attendees honed in on one particular challenging problem for future AR/MR technology designs: *how to manage the interaction of virtual content with the ownership of physical space.* In this section, we describe the outcomes of that discussion. In particular, we do not seek to provide prescriptive technology designs, but rather to lay out possible harms and design considerations that must be taken into account alongside an AR/MR platform or application developer's functionality and other goals.**

The core setting we consider is the following: an AR/MR platform or application creates virtual content and displays it overlaid on top of locations and objects in the physical world. Depending on the design of the platform, there may be different ways that the world is mapped and that content is discovered by users. For example, users might subscribe to certain content "channels" (such as a "Wikipedia" channel that gives localized information in an AR display about the user's surroundings) or see content only from applications that they choose to install and run (such as a Pokémon application). Or users may see only content that has been explicitly shared with them by other users, or only content that they themselves have created. Regardless of how virtual content is linked to the physical world, and how it comes to be seen by a given user, a key question arises: what can go wrong when virtual content is overlaid on a physical space, and visible to at least some users?



Stated in terms of the framework described in the previous section, we considered the broad question: *What happens when an AR/MR platform or application displays virtual content to one or more users overlaid onto their views of the physical world in different settings?* Exploring this question surfaced a wide range of different concerns that users or bystanders of an AR/MR platform or application might have around where and what virtual content is placed in the views of some or all users. We emphasize that a given AR/MR platform design may or may not actually enable each of the following potential harms, but that it is crucial to consider whether and how they might arise or be mitigated.

### Potential concerns users may have about virtual content they or others see overlaid on the physical world:

UNDESIRABLE CONTENT:

- Will I see ads plastered onto everything?
- Will people see content that is inappropriate for the context (e.g., content seen by children)?
- Will I see content that I find startling or disturbing or harassing as I move through the physical world? What will I be able to do about it?

UNDESIRABLE CONTENT PLACEMENT:

- Will virtual content block real world objects in disruptive ways?
- Will virtual content from multiple sources interfere with each other?
- Will other users be able to place virtual content (that I can or cannot see) overlaid on my body (e.g., an offensive hat)? Will other users be able to modify my appearance even more fundamentally in their (and/or others') view of the world (e.g., making me appear naked, or as someone completely different)?
- Will other users be able to place virtual content (that I can or cannot see) in locations that I own (e.g., my house, my business), on my physical objects (e.g., my handbag, my car), or on my virtual objects (e.g., my virtual pet)?
- Will other users or AR/MR applications be able to modify my trademarked content in the physical world (e.g., the design of a Coke can)?
- Will I be able to know or control whether and what virtual content is overlaid on my body or on physical spaces that I own or control?

ACCESS TO CONTENT:

- Who will be able to see virtual content I create in different places? How will I know?
- Will others be able to use my likeness, or content that I created, without my permission and/or without me knowing?
- What platforms will I be able to use to access my content? Will I be locked into a content ecosystem that is separate from other users?

ATTRIBUTION:

- Will I be able to know what caused the bad content that I saw (e.g., so I can take an action like uninstalling an application, unsubscribing from the source, or blocking a problematic user)?

FUNCTIONALITY LIMITATIONS:

- Will I be able to do what I want with my AR/MR device?
- Will restrictions interfere with my need to use AR/MR for assistive purposes (e.g., translation, subtitling)?

## DESIGN CONSIDERATIONS

We next break down a number of different considerations when determining whether and how to address each of the above concerns in the design of an AR/MR platform or application.

### Roles of different stakeholders

Challenges and solutions will incorporate a variety of different stakeholders; designers should consider not only how these stakeholders might be impacted, but also their potential roles in a solution. These stakeholders include platform developers, app developers, content creators or moderators, users, bystanders, physical space owners, and rule and policy makers (from local to global levels).

### Taxonomy of spaces

Different norms, regulations, or policies may apply or be appropriate solutions in different types of spaces. Consider:

- Public spaces (e.g., parks or roads), semi-public spaces (e.g., a shopping mall), private spaces (e.g., a person's home), corporate spaces (e.g., a company office), government spaces, private spaces that can be seen from public vantage points (e.g., a front lawn).
- Absolute spaces (i.e., spaces referenced by particular coordinates in the physical world) versus relative spaces (i.e., spaces referenced in relation to a moving person, animal, or object in the real world, or other virtual objects).
- Physical versus virtual spaces.

### Strawman and challenges

As one model, one might consider hierarchies of different types of spaces. For example, a conference room inside a university building exists inside nested jurisdictions, e.g., Room – University – City – State – Country. If the university has certain policies about AR content, then those policies might equally apply inside the room. However, while this hierarchical model presents some potential solutions, it also raises challenging questions:

- How do absolute and relative space interact? For example, if a person walks through a public space, how should that person's "don't display AR content on my body" interact with the space's default "any content allowed anywhere" policy? One possible approach is to take the most restrictive intersection of all policies for a given point in physical space at a given moment in time.
- However, what should be done when the policies of the different hierarchies conflict in ways that the most-restrictive intersection is not appropriate? For example, consider the case where the meeting in the conference room is concerned specifically with research about the type of content that is prohibited by the university policy. Or if an AR/MR device being is being used as an assistive technology, it is likely desirable for this to override local policies — but who determines what is treated as an assistive device?
- More generally, who determines the jurisdictions in the hierarchy, and how is space ownership determined such that the appropriate person or people are able to set the policy on the space? How does a device or application determine and authenticate the relevant policies in a given space? The challenges here include not only identifying who is the legal owner of a space and what policies apply, but also how to handle more ambiguous ownership situations (e.g., does the landlord or the person living in a rented house have priority?).

### Design axes

Different AR/MR platforms and applications should and will undoubtedly differ in how (and whether) they choose to answer the above questions. Again, our goal in this section is not to prescribe a particular solution or design. Towards that end, we end by presenting a set of design axes, informed by the above concerns and challenges, that platform and application designers must consider:

- *How and by whom is virtual content created?* For example, by multiple different application developers or content creators, by users themselves, or only centrally by the application or platform?
- *How do users come to see virtual content?* For example, do they seek it out explicitly? Is it shared with them explicitly by other users? And/or do they see it automatically when they walk into certain physical spaces?
- *How and by whom is content placed?* For example, by applications, and/or explicitly by users? What is the platform's role in managing content placement? Is content from multiple sources or applications shown at once? Prior academic work has begun studying these issues (Lebeck et al. 2017, Lebeck et al. 2019).

- *Who determines where content is or can be placed, relative to the physical world?* The platform or application? The content creator? People who own fixed physical spaces (e.g., land)? People who own moving objects or their own bodies? The users who view the content? Some combination of the above?
- *If external entities (e.g., land owners) create policies on where content can be placed, how are those policies communicated and ownership verified?* For example, policies might be centrally registered with the AR/MR application or platform in question, and physical location ownership might be verified by sending physical postcards (as is done by Google Maps[3], though work by Huang et al. showed how this method has been circumvented in practice). Related academic work has proposed privacy policies for AR devices based on signals from the physical world ("world-driven access control", Roesner et al. 2014).
- *Is policy enforcement proactive or reactive?* That is, can/will the platform or application prevent content from being created and/or placed in the world, and/or is problematic content taken down after the fact (e.g., after multiple users report the content)?
- *How should different types of content be treated differently?* For example, a given stakeholder may have different preferences or policies for content that is not appropriate for children.
- *How (if at all) will authentication and attribution be handled?* For example, can content that harasses a user be attributed (by that user and/or by the application/platform in question) to the user who created and/or placed it?

[3] https://support.google.com/business/answer/4588357?hl=en

5

# CONCLUSION AND NEXT STEPS

This report has asked many questions and provided few solid answers. It is clear that substantial open questions and key challenges remain towards achieving security, privacy, and safety in AR/MR technologies while reaping their full benefits. These are crucial issues: if they are not addressed, then they pose direct risks to the consumers of these technologies as well as an existential threat to the widespread adoption of these technologies themselves, undermining their potential positive impacts on individuals and society. In summarizing the observations from a two-day Summit among industry and academic leaders in the space of AR/MR and computer security & privacy, we have aimed to draw attention to these issues — in general and in specifics — as well as to lay a foundation for a path forward.

This report is a call to action: We call on AR/MR designers use the general framework provided in this report to consider the potential impacts of the technologies they are building and to make *conscious* choices about tradeoffs when a perfect solution is impossible. We call on technology creators and researchers to develop solutions to the specific challenges we raised. We call on policymakers to understand these issues and to consider the potential role of policy and regulation in the solution space (an issue we did not have the expertise to discuss in depth at the Summit), but to do so responsibly and not prematurely, and with the consultation of technology experts. Finally, in our view, one of the greatest successes of our Summit was to bring together expertise and perspectives from different backgrounds and different industry players into one unified conversation, stepping back from the specific goals of any individual company or stakeholder. We all share the same overarching goal, and we call on these diverse sets of stakeholders to continue to discuss and work together towards a future in which AR/MR technologies fulfill their positive potential while protecting the security, privacy, and safety of their users and others.

6

# REFERENCES AND ADDITIONAL READINGS


**S. Ahn, M. Gorlatova, P. Naghizadeh, M. Chiang, P. Mittal**. "Adaptive Fog-based Output Security for Augmented Reality." In *Proceedings of the ACM SIGCOMM VR/AR Network Workshop*, August 2018. https://maria.gorlatova.com/wp-content/uploads/2018/06/AdaptiveFogBasedSecuritySAhn2018.pdf.

**S. Baldassi, T. Kohno, F. Roesner, and M. Tian.** "Challenges and New Directions in Augmented Reality, Computer Security, and Neuroscience — Part 1: Risks to Sensation and Perception." arXiv:1806.10557, June 2018. https://arxiv.org/abs/1806.10557.

**De Guzman, J.A., Thilakarathna, K., and Seneviratne, A.** "Security and Privacy Approaches in Mixed Reality: A Literature Survey." In *ACM Computing Surveys*, Volume 52, Issue 6, Article 110, October 2019. https://arxiv.org/abs/1802.05797

**Denning, T., Dehlawi, Z., and Kohno, T.** "In Situ with Bystanders of Augmented Reality Glasses: Perspectives on Recording and Privacy-Mediating Technologies." In *Proceedings of the Conference on Human Factors in Computing Systems (CHI),* April 2014. https://ar-sec.cs.washington.edu/files/ar-chi2014.pdf.

**Hosfelt, D.** "Making ethical decisions for the immersive web." Mozilla Mixed Reality Blog, May 2019. https://blog.mozvr.com/making-ethical-decisions/.

**Huang, D. Y., Grundman, D., Thomas, K., Kumar, A., Bursztein, E., Levchenko, K., and Snoeren, A. C.** "Pinning Down Abuse on Google Maps." In *Proceedings of the International Conference on World Wide Web (WWW)*, 2017. https://research.google/pubs/pub45976/.

**Jang, Y., Song, C., Chung, S., Wang, T., and Lee, W.** "A11y Attacks: Exploiting Accessibility in Operating Systems." In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*, November 2014. http://wenke.gtisc.gatech.edu/papers/a11y.pdf.

**John, B., Koppal, S., Joerg, S., Jain, E.** "The Security-Utility Trade-off for Iris Authentication and Eye Animation for Social Virtual Avatars." In *IEEE Transactions on Visualization and Computer Graphics* (TVCG 2020), Special Issue on IEEE VR.

**Lebeck, K., Kohno, T., and Roesner, F.** "Enabling Multiple Applications to Simultaneously Augment Reality: Challenges and Directions." In *Proceedings of the 20th Workshop on Mobile Computing Systems and Applications (HotMobile),* February 2019. https://ar-sec.cs.washington.edu/files/lebeck-arsec-hotmobile19.pdf.

**Lebeck, K., Ruth, K., Kohno, T., and Roesner, F.** "Securing Augmented Reality Output." In *Proceedings of the 38th IEEE Symposium on Security and Privacy (Oakland),* May 2017. https://ar-sec.cs.washington.edu/files/lebeck-sp17.pdf.

**Lebeck, K., Ruth, K., Kohno, T., and Roesner, F.** "Towards Security and Privacy for Multi-User Augmented Reality: Foundations with End Users." In *Proceedings of the 39th IEEE Symposium on Security and Privacy (Oakland),* May 2018. https://ar-sec.cs.washington.edu/files/arsec-lebeck-sp18.pdf.

**Liu, A., Xia, L., Duchowski, A., Bailey, R., Holmqvist, K., and Jain, E.** "Differential Privacy for EyeTracking Data." In *Proceedings of ACM Symposium on Eye Tracking Research & Applications* (ETRA'19).

**Raval, N., Srivastava, A., Razeen, A., Lebeck, K., Machanavajjhala, A., and Cox, L. P.** "What You Mark is What Apps See." In *Proceedings of the 14th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, June 2016. https://users.cs.duke.edu/~lpcox/raval-mobisys16.pdf.

**Roesner, F., Kohno, T., and Molnar, D.** "Security and Privacy for Augmented Reality Systems." In *Communications of the ACM,* Vol. 57, No. 4, Pages 88-96, April 2014. https://ar-sec.cs.washington.edu/files/arsec-cacm2014-preprint.pdf.

**Roesner, F., Kohno, T., Moshchuk, A., Parno, B., Wang, H. J., Cowan, C.** "User-Driven Access Control: Rethinking Permission Granting in Modern Operating Systems." In *Proceedings of the 33rd IEEE Symposium on Security and Privacy (Oakland)*, May 2012. https://www.franziroesner.com/pdf/udac-oakland2012.pdf.

**Roesner, F., Molnar, D., Moshchuk, A., Kohno, T., and Wang, H. J.** "World-Driven Access Control for Continuous Sensing." In *Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS),* November 2014. https://ar-sec.cs.washington.edu/files/wdac-ccs2014.pdf.

7

# NAMED ATTENDEES

**Ruth, K., Kohno, T., and Roesner, F.** "Secure Multi-User Content Sharing for Augmented Reality Applications." In *Proceedings of the 28th USENIX Security Symposium,* August 2019. https://ar-sec.cs.washington.edu/files/ruth-arsharing-usenixsec19.pdf. ShareAR software toolkit: https://www.arsharingtoolkit.com.

**Speicher, M., Hall, B.D. and Nebeling, M.** "What is Mixed Reality?" In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 537). ACM, April 2019. http://michael-nebeling.de/publications/chi19b.pdf.

**Steil, J., Hagestedt, I., Huang, M. X., and Bulling, A.** "Privacy-aware eye tracking using differential privacy." In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 2019.

**Steil, J., Koelle, M., Heuten, W., Boll, S., and Bulling, A.** "Privaceye: Privacy-preserving head-mounted eye tracking using egocentric scene image and eye movement features." In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 2019.

**Our Summit included participants from both academia and industry. These included the following people, who gave permission to be named in this report:**

Avi Bar-Zeev, RealityPrime

Linda Ng Boyle, University of Washington

Neil Coles, Microsoft

Shaheen Gandhi, Facebook

Forest Gibson, Pluto VR

Maria Gorlatova, Duke University

Diane Hosfelt, Mozilla

Eakta Jain, University of Florida

Tadayoshi Kohno, University of Washington

Mark Lucovsky, Facebook

Blair MacIntyre, Mozilla and Georgia Tech

Kate McKinley, Facebook

Justin Quimby, Google

Franziska Roesner, University of Washington

Kimberly Ruth, University of Washington