

SHOULДАР: Detecting Shoulder Surfing Attacks Using Multimodal Eye Tracking and Augmented Reality

MATTHEW CORBETT, Virginia Tech, USA

BRENDAN DAVID-JOHN, Virginia Tech, USA

JIACHENG SHANG, Montclair State University, USA

BO JI, Virginia Tech, USA

Shoulder surfing attacks (SSAs) are a type of observation attack designed to illicitly gather sensitive data from “over the shoulder” of victims. This attack can be directed at mobile devices, desktop screens, Personal Identification Number (PIN) pads at an Automated Teller Machine (ATM), or written text. Existing solutions are generally focused on authentication techniques (e.g., logins) and are limited to specific attack scenarios (e.g., mobile devices or PIN Pads). We present SHOULДАР, a mobile and usable system to detect SSAs using multimodal eye gaze information (i.e., from both the potential attacker and victim). SHOULДАР uses an augmented reality headset as a platform to incorporate user eye gaze tracking, rear-facing image collection and eye gaze analysis, and user notification of potential attacks. In a 24-participant study, we show that the prototype is capable of detecting 87.28% of SSAs against both physical and digital targets, a two-fold improvement on the baseline solution using a rear-facing mirror, a widely used solution to the SSA problem. The SHOULДАР approach provides an AR-based, active SSA defense that applies to both digital and physical information entry in sensitive environments.

CCS Concepts: • **Security and privacy** → **Usability in security and privacy**; **Social aspects of security and privacy**; **Usability in security and privacy**; **Social aspects of security and privacy**; • **Human-centered computing** → *Mobile devices*; *Mobile devices*.

Additional Key Words and Phrases: Security, Privacy, Eye Gaze, Augmented Reality

ACM Reference Format:

Matthew Corbett, Brendan David-John, Jiacheng Shang, and Bo Ji. 2024. SHOULДАР: Detecting Shoulder Surfing Attacks Using Multimodal Eye Tracking and Augmented Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 97 (September 2024), 23 pages. <https://doi.org/10.1145/3678573>

1 Introduction

Shoulder surfing attacks (SSAs) are a type of observation attack designed to gather sensitive information (e.g., passwords, social security numbers, or home addresses) from the victim by directly observing this information without the victim’s knowledge. This is generally done “over the shoulder” of the victim in a way that does not allow the victim to notice this attack, as illustrated in Fig. 1. Published work on this topic is plentiful, with 8,420+ works related to “shoulder surfing” listed on a recent (February 2024) Google Scholar search [28], including 3600 being published since 2017 [24].

The vast majority of research on SSAs has focused on reducing SSAs *during authentication* (e.g., while entering passwords) through advanced authentication schemes or other obfuscation methods, with 70 of the 92 works published in the last year focused on improved authentication schemes [53]. These works require biometrics,

Authors’ Contact Information: [Matthew Corbett](#), matthewc84@vt.edu, Virginia Tech, Blacksburg, Virginia, USA; [Brendan David-John](#), bmdj@vt.edu, Virginia Tech, Blacksburg, Virginia, USA; [Jiacheng Shang](#), shangj@montclair.edu, Montclair State University, Montclair, New Jersey, USA; [Bo Ji](#), boji@vt.edu, Virginia Tech, Blacksburg, Virginia, USA.



This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2024 Copyright held by the owner/author(s).

ACM 2474-9567/2024/9-ART97

<https://doi.org/10.1145/3678573>



Fig. 1. An illustration of two scenarios of SSAs: the attacker is attempting to observe (a) sensitive information on a victim's laptop and (b) the written information of the victim. Images were generated using Microsoft Designer [49].

graphical passwords, or other advanced input mechanisms to prevent an attacker from seeing sensitive authentication information. These solutions have also been shown to introduce extra time and effort for authentication [23]. Other solutions obscure the attacker's view of the sensitive information [68]. We refer to these solutions as *passive solutions*, as none of these works give the victim an indication that an attack is in progress. Not only do these solutions provide little protection for other information (e.g., their written addresses, bank account information, video chats, and texts), but they also offer no protection *after the user has completed the authentication*.

Active solutions detect and alert the user to a potential SSA in real time. This allows the user to react and reduce or remove the threat. The most recent active solution is nearly two years old and is limited to mobile devices [10]. Of the remaining three known active solutions, two are designed only for mobile devices, and the other is a convex mirror normally attached to ATMs [4, 51, 62]. The small number and limited scope of existing solutions show that more research into these solutions could be fruitful in improving existing anti-SSA techniques.

While both types of solutions differ in their protection methods, passive and active solutions are currently limited to a specific type of protection domain. These solutions work for either mobile devices, Personal Identification Number (PIN) entry pads at an Automated Teller Machine (ATM), desktop/laptop computers, or other devices, and one solution does not protect users across multiple domains. This limits each solution's real-world feasibility and makes them less usable.

In this work, we define *usable* in two parts: the ability of the solution to protect the individual user in a wide variety of scenarios, and the solution's minimal negative impact on the user's completion of the potentially SSA-vulnerable task. To highlight the lack of usable solutions to SSAs, we offer a simple observation. Throughout daily life, especially since the advent of mobile devices, people worldwide are often asked to write and type sensitive information regularly. It is easy to imagine a person being asked to write an address or personal information on a government form and log into a bank account/ATM on the same day. As of the time of writing this paper, no existing solution could protect individuals from SSAs during all of these activities as they occur in different environments and scenarios. Additionally, most existing solutions are designed to only protect information during authentication, such as during PIN or password input, and add extra time and effort to this process.

To create a unified solution to these different attack scenarios, we must understand the basic components of any SSA. With few exceptions, these attacks involve a victim using their eyes while entering or writing sensitive

information and an attacker using their eyes to observe it. Exceptions, such as electronic surveillance, exist but are generally not as prominent [6, 7, 11, 24, 35]. Using this understanding, we can imagine a solution that uses multimodal eye tracking (i.e., from both the attacker and victim) as a universal method to detect and warn of SSAs. Eye gaze from the user can be used to identify the user’s focus and can also be used to determine when the user is concentrating and vulnerable to SSAs. Tracking the eyes of persons to the user’s rear can identify when these persons are looking over the shoulder of the user. Head-mounted eye-tracking devices provide the ability to track the eyes of the device user to monitor their focus through information such as eye fixations [44, 50]. Finally, with the addition of a modular rear-facing camera and readily available facial and eye landmark detection methods, we can analyze the potential attacker’s gaze direction.

Such a solution also needs to notify the user of a detected attack. To address a potential attack, the device user must know generally where the attack is occurring to protect themselves. If the user does not know this, moving or adjusting the position of their sensitive information may not have the desired effect. Head-mounted devices, such as augmented reality (AR) headsets, discreetly provide this capability, only visible to the user. A recent article from the National Institute of Standards and Technology (NIST) recommended AR to “help preserve data confidentiality while displaying sensitive information compared to handheld or desktop displays” [26]. While the article directly refers to rendering information so only the user can see it, we imagine leveraging the suite of sensors integrated within AR devices as a means to achieve data confidentiality that is applicable within physical mediums. Currently, AR headsets are not as ubiquitous as other devices, like mobile phones, but are expected to become more and more common as they become lighter and less obtrusive [43, 46]. For example, the Apple Vision Pro is pitched as a daily productivity tool to replace or enhance a traditional office computing setup. Sales of the Vision Pro are expected to exceed 400,000 units in 2024 and signals an increase in consumer adoption of AR headsets [9, 29].

AR headsets are an ideal platform for our design, as these devices often contain eye tracking, provide a method to alert the user through the device’s display, and are mobile [67]. For these reasons, we will adopt head-mounted AR to harness user eye gaze, collect rear-facing camera frames (with an added off-the-shelf camera), and discreetly notify the user of an attack. Light-weight AR headsets can be worn in various tasks and environments, including common SSA attack scenarios (typing, writing, using an ATM, etc.) [43]. We adopt AR headsets with a suite of data sensors for our ideal system design, though not all smartglasses fit the currently proposed design.

We summarize the contributions of this paper as follows:

- We present our design called SHOULДАР, a mobile and usable system designed to protect users from SSAs. Using an off-the-shelf camera and microcontroller, and an AR headset, we leverage multimodal eye tracking (i.e., eye gaze information from both the user and attacker) to alert the device user to potential shoulder surfing threats in both typing and writing scenarios.
- Through a user study with 24 participants, we evaluate the effectiveness of this system, the effect of the system on comfort in SSA-vulnerable situations, and the perceived usability. Our solution successfully detected an SSA in 87.28% of tests across *both* typing and writing tasks, a more than two-fold increase in protection over the baseline solution using a rear-facing mirror, a widely used solution to the SSA problem.
- Our system improves perceived participant comfort in various representative scenarios and was considered more usable than a baseline solution.

2 Related Work

We begin with a brief review of existing work on the topic of SSA. While not exhaustive, we believe that the review is representative of the state-of-the-art. We divide this section into three parts: a discussion about the background of SSA, a review of passive solutions (which do not indicate ongoing attacks on the victim), and active solutions (which do).

2.1 Shoulder Surfing Attacks

A shoulder surfing attack (SSA) is a type of observation attack designed to gather sensitive information about the victim, such as authentication information or personally identifiable data. This attack occurs in public situations (e.g., public transportation or an office building) [36]. While most reported attacks are against mobile device users, attacks can be directed at various targets. SSAs can be on mobile electronic devices (e.g., cell phones), PIN entry pads at an ATM [11], or sensitive written information like Social Security Numbers or home addresses [59]. These attacks allow the attacker to obtain sensitive information that can provide insight into the victim's hobbies, relationships, sexual preferences, finances, and login data [24]. Fig. 1 gives an illustration of two of these attack scenarios. These attacks are generally not noticed by the victim, with only 7% of attacks noticed in a controlled study involving attacks on mobile devices [24]. While this shows that SSAs are generally unnoticed, 90.2% of participants in a survey studying SSAs reported that they would feel "anger, unease, pressure, and embarrassment" upon noticing an SSA directed at their information [24].

While simply observing sensitive information over the shoulder of a person is not new, the vast majority of research on SSAs had been published since the announcement of the original iPhone in 2007 [24]. Given the focus on mobile devices, existing techniques to address SSAs generally focus on protecting the users of mobile phones, although some solutions do exist to protect other attack surfaces, like ATM PIN pads. For example, local regulations in some U.S. states require the installation of a small fisheye convex mirror to allow rear-facing views when entering PINs and retrieving cash [41]. We divide existing solutions into two categories: passive and active mitigation techniques. Simply put, passive techniques rely on obfuscating or complicating the shoulder-surfing attacker's attempt to collect the victim's sensitive data, not detecting the attack while it is in progress. Active techniques seek to identify the attacker in real-time and alert the victim, giving them a chance to intervene, move, or otherwise address the situation. Passive techniques prevent attacks, while active techniques detect the attack as it happens.

2.2 Passive Solutions

Passive techniques are those that do not seek to identify the attacker or react to the attack; they merely seek to make the chances of the attacker's success lower. Existing passive SSA solutions can be divided into five categories, based on the methods used to defeat the potential attacker. (1) Eye-gaze-based techniques let users use their eye gaze to enter sensitive information, such as on a PIN pad [22] or on a keyboard [37] and make the information more difficult to observe. Additional eye-gaze-based techniques use "eye gestures" or specific eye movements for authentication or data entry. (2) PIN-based methods create PIN entry schemes designed to make it difficult for the attacker to observe, record, or guess the PIN [12, 21, 38, 57, 61]. These techniques seek to reduce attack success by altering the PIN entry format and complicating the attacker's ability to observe. (3) Pattern, location, or graphic-based techniques seek to remove a uniquely identifying item such as a password or PIN and add in a pattern like a series of pictures, or physical location only an authorized user would know [14, 15, 17, 23, 30]. (4) Obscuration-based methods provide a physical barrier to SSAs (e.g., a screen protector) such that the information can only be viewed from directly in front of the device [5, 16, 68]. Some techniques are designed for users without the ability to use their eyes, such as the visually impaired, and use vibration or Braille-based methods [6, 7]. Finally, (5) multimodal techniques seek to combine eye gaze and physical motion to overcome the ability of attackers to track user eye gaze and simulate input [3, 32, 33]. These techniques comprise the majority of existing solutions to SSAs, but they are not usable in scenarios such as protecting sensitive information within written documents.

Active SSA Solutions	Protection Domain	
	Digital	Physical
PrivacyScout [10]	✓	✗
iAlert [4]	✓	✗
Are They Actually Looking? [62]	✓	✗
ATM Mirrors [41]	✗	✓
SHOULДАР (this work)	✓	✓

Table 1. A comparison of current active solutions to SSAs.

2.3 Active Solutions

Active SSA solutions seek to detect an attack in real time and to potentially prevent the attack from ever occurring. Table 1 outlines existing active solutions and separates them by their protection domain. Some techniques seek to alert a user to a potential attacker by identifying faces in the field-of-view of a front-facing camera and analyzing eye gaze [4, 10, 62]. Other techniques use mirrors, like the built-in protections at ATMs, to help the potential victim identify these faces [41, 51]. These techniques all rely on locating faces around the user and evaluating the distance of the face and/or the person's eye gaze to determine the attack likelihood. However, these solutions are limited by the field of view of the device's camera or mirror. If the user is positioned in a way that blocks the camera or mirror's view of the attacker, it will be hard to detect the attack. Also, these solutions are limited to a single domain, such as protecting a PIN pad or a mobile device. With the potential exception of the ATM PIN Pad, no current active solution can detect or prevent attacks in other scenarios. Finally, these solutions are generally not mobile. An ATM mirror can protect the user at the ATM or in its vicinity, but cannot protect this person once they move away. Existing active solutions, like passive techniques, lack broad use cases.

Passive and active solutions are convenient ways to divide existing work but are also orthogonal to each other. Because passive solutions protect authentication, and active solutions alert the user to SSAs in specific scenarios, we cannot compare them directly. We present these solutions not in direct comparison but as an example of the lack of usability in existing work.

3 Challenges and Key Insights

We seek a solution that improves the usability of existing work based on the definition presented in Section 1. If we imagine such a solution, two key requirements begin to emerge: **(R1)** the system must be generalizable to a myriad of scenarios (e.g., ATMs, office desks, whiteboards, mobile devices, and desktop/laptop computers); **(R2)** the system must not interfere with the user's actions. Specifically, the system cannot prohibit typing and writing or create long task completion times (e.g., login times). As passive systems generally only protect authentication and are not generalizable, we seek an active solution to this problem. Next, we explore two prominent challenges of such an active system.

3.1 Challenges

A usable solution should work in various types of scenarios. While Section 2 reviews multiple works aiming at deterring or identifying SSAs, each solution is confined to a particular domain. For instance, graphical or PIN-based methods require a digital display or a physical PIN pad. Existing works designed to alert users that a person is looking at their device generally require front-facing cameras that can be obscured by the user and are not generalizable solutions. Rear-facing mirrors are mostly mounted on ATMs and are generally not mounted in other areas, like in office settings or on desktop/laptop computers. Therefore, *the first challenge is to create a method that can be generalized across multiple domains, both digital and physical (C1).*

A usable solution must not complicate the user's task or add cognitive burden. Passive techniques, like those discussed in Section 2, have been shown to require good color vision or hand-eye coordination and can be difficult for the elderly or persons with intellectual disability to use [13]. Additionally, many of these schemes have been shown to increase task completion (e.g., login) times or add a cognitive burden to users that have been deemed unacceptable [23].

Some existing active solutions, like the ATM mirror, require the user to constantly monitor their surrounding for threats. Such a requirement can add a cognitive burden to the user by forcing the user to assess and respond to perceived attacks on their sensitive information. Additionally, to ensure that an active solution is timely and useful, the system should operate in real time. If not, the user would have to decide if the threat is still present. This can remove attention from the task at hand. Therefore, *the second challenge is to create a system that runs in real-time, does not interfere with task completion, and does not place a cognitive burden on the user (C2).*

3.2 Key Insights

Eye gaze can provide an avenue for a generalizable and effective solution. To define the common denominators of SSAs, we distill the SSA problem into its core components. An SSA is, generally, (1) an attacker using their eyes to observe sensitive information, and (2) the victim using their eyes to simultaneously attend to and/or reveal sensitive content. While not every situation involves the attacker or the victim's eyes, such as in situations where the potential victim has a visual impairment or the attacker uses electronic surveillance, this definition covers the majority of attack scenarios [6, 7, 11, 24, 35]. Additionally, this definition unifies nearly all attack targets, including observing the victim's written information, ATM PIN, and information typed on laptops, desktops, and mobile devices.

Understanding the importance of gaze and attention to SSAs, we then leverage works from the field of psychology to extract salient features of the eye that can aid in our efforts to find an effective solution. For attackers, we can use eye gaze to create a more fine-grained estimation of their target and match their eye gaze to that of the user. This technique can remove instances where the potential attacker's head may be directed at something innocuous, but their actual eye gaze is directed at sensitive information. Additionally, this prevents false positive indications from the opposite scenario.

For the user, eye gaze information can provide fine-grained details about where the user is looking, how hard the user is concentrating on a given task or object, the user's emotions, and more [2, 50, 52]. Assuming that increased concentration on a target, say while writing or typing a password, creates a decreased lack of awareness of the user's surroundings, we can also use this metric to gauge the potential vulnerability of the victim to SSAs. If we use a proper threshold for this concentration, we can then decide to alert the user to SSAs only at times when they are less capable of detecting the attack themselves. Existing work gives specific thresholds for fixation duration, the amount of time an eye dwells on a particular area, which can be used as a metric for concentration and attention [50, 52, 63, 72]. Thus, *eye gaze is a nearly universal component of SSAs*. We can leverage this fact to create a solution that constantly evaluates the eye gaze of the potential victim and attacker and overcome Challenge (C1).

Augmented reality (AR) headsets and eye tracking, together, create an opportunity for a usable solution. As the human eye is at the root of the SSA problem, we seek to tie a viable platform to this capability. Most modern AR headsets are equipped with eye-tracking sensors capable of determining user eye gaze direction and other features of the human eye such as saccades, fixations, and pupil dilation [9, 40, 71]. Given the previous insights about the usefulness of eye gaze, we can leverage the onboard eye tracking of these devices as a platform to overcome Challenge (C1). We can also use the inherent functionality of AR headsets (e.g., the ability of the AR user to interact with the physical environment while wearing the device) to make the solution mobile and more usable. AR headsets are designed to allow the user a view of the physical world with an overlay of digital content.

This allows the user to interact with the physical world as they normally would. As an example, a user can access their bank accounts at an ATM, use a laptop or desktop, or write with pen and paper all while wearing an AR headset. While current AR headsets are wearable and mobile, their weight and battery life limit their utility. Future AR devices, however, are not expected to have this problem [43, 46]. As these headsets get smaller, lighter, and last longer between charges, they may become as common as cell phones.

Additionally, the integrated onboard processors and graphics processing units (GPUs) give AR devices adequate computing power to accomplish our required facial detection and gaze estimation tasks. Finally, assuming the AR-based system only announces an attack when detected and does not require the user to interact otherwise, the notification of the attack is the only expected cognitive load. AR headsets give us the computing power and mobility to address Challenge (C2).

With these insights, we can now imagine a mobile, usable, active system using eye gaze and AR that overcomes the challenges presented in Section 3.1 and identify research questions to be answered as part of our experimentation, namely:

- (1) Can we use insights into human eye gaze to create a usable, AR-based, anti-SSA system?
- (2) Can such a system reduce participant discomfort in SSA-vulnerable situations?
- (3) Does this system reduce discomfort more effectively than existing solutions?

4 SHOULДАР Design

This section incorporates the insights in Section 3 into a coherent design, called SHOULДАР, centered around eye tracking and AR. SHOULДАР is based on an AR headset that includes eye-tracking sensors and is capable of detecting attacks in a wide variety of scenarios. Additionally, all the information collection, processing, and decision-making are transparent to the user and conducted implicitly by the device and in real time. The user simply has to wear their AR headset while an application such as SHOULДАР is running, and does not need to remember specialized information, such as PIN or image combinations. By combining these advantages, the SHOULДАР design presented in this section can protect against SSAs in a wide variety of scenarios. We begin with a brief review of the threat model with an eye toward the specific limitations of a feasible SSA, followed by discussions about the key components of SHOULДАР.

4.1 Threat Model

Informed by existing work [10, 31, 57], we define the limits of the SSA in terms of space and angle from the target of the attack. SSAs are generally from behind the user, where the user cannot detect and prevent the attack. While attacks are possible against large displays at larger distances, existing literature on the topic shows that a maximum attacker distance of 2 meters is generally the farthest an attack can feasibly be conducted against a wide range of targets (mobile devices, paper, desktops, etc.) [10]. Also, the field of view (FOV) of the average human (assuming a stationary head, but mobile eyes) is as much as 290° [31]. This is also known as the *total visual field*. This leaves roughly 70° of FOV where the person cannot detect objects or motion to their rear. We call this area roughly 70° behind the user at no more than 2 meters the *vulnerable area*. Any SSA from outside the vulnerable area would risk the possibility of the potential victim noticing this attack.

For this reason, we limit our threat model to 40° on either side of the potential victim (a total of 80°, which covers the vulnerable area) as measured from directly behind the person, allowing for a small margin at the extreme edges of human sight. Fig. 2 gives an illustration of this area. We acknowledge the possibility that the user may be concentrating on an area near the center of the FOV and would be unlikely to move their eyes left or right in this scenario. This would increase the size of the vulnerable area. We also note that a single fixation at the center of their FOV is temporary. Humans make roughly two to four rapid gaze shifts (saccades) per second [39], providing an opportunity to visually scan their periphery in response to new faces or bystanders. This would

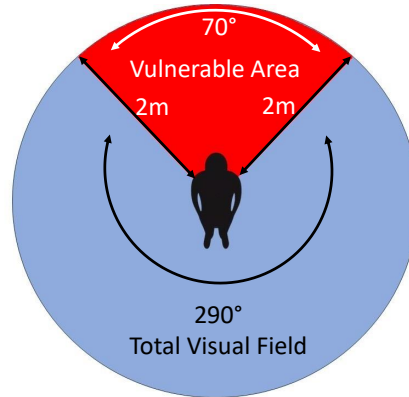


Fig. 2. SHOULDAR Threat Model. Given the human eye’s ability to sense motion at nearly 290° when allowed to move the head and eyes (with the torso stationary), we consider the remaining 70° to be the most vulnerable area for an SSA. For evaluation, we expand our “vulnerable area” to 80° to account for attackers on the edge of this area. Additionally, previous work shows that attacks on mobile devices and other common targets are most feasible inside 2 meters. We define the *vulnerable area* in red as the most likely area for an SSA. However, this vulnerable area could be larger if the victim was concentrating on an area at the center of their field of view, limiting their ability to move their eyes left and right. We discuss this in greater detail in Section 7.

reduce the impact of focused vision on SSA awareness due to our ability to perceive the presence of faces in the periphery [56]. We discuss this limitation and present potential solutions in Section 7. Finally, in this threat model, we assume that the face of the potential attacker is visible and is not deliberately obscured by either digital facial obscuration method such as [58], or any other method such as using a hand or an object. We discuss this more in Section 7 as well.

4.2 AR Headsets - A Feasible Platform for Generalized Anti-SSA Techniques

As noted in Section 3.2, AR headsets provide a suitable platform for our envisioned SSA detection technique. AR headsets are portable and can be expected to be worn in a wide variety of situations, especially when considering the lighter and less cumbersome versions expected to become a reality [60]. These future devices are even hypothesized to replace mobile phones at some point in the future [46]. Additionally, AR headsets normally have eye-tracking sensors that give us insight into various attack scenarios. Our design centers around such an AR headset. However, to observe the area behind the user (i.e., the vulnerable area), we require a rear-facing camera that most current AR headsets do *not* have. To overcome this obstacle, we retrofit the headset with a rear-facing camera, which can be integrated into future headsets. With this addition, we can simultaneously estimate the eye gaze of the potential attacker and user on a mobile device usable in various scenarios.

4.3 Design Components

Fig. 3 illustrates the SHOULDAR design. SHOULDAR collects raw eye tracking and camera data (Step 1), analyzes this data (Step 2), evaluates user vulnerability and threat (Step 3), and alerts the user if required (Step 4).

Step 1: Data Collection. Our design must collect multimodal eye gaze information about the user and potential attackers in the physical environment, specifically in the vulnerable area up to 2 meters and in a 70° area behind the user. With the retrofitted, rear-facing camera, we can capture images directly behind the device user and identify potential attackers. Additionally, using the inherent functionality of the AR headset, we can capture

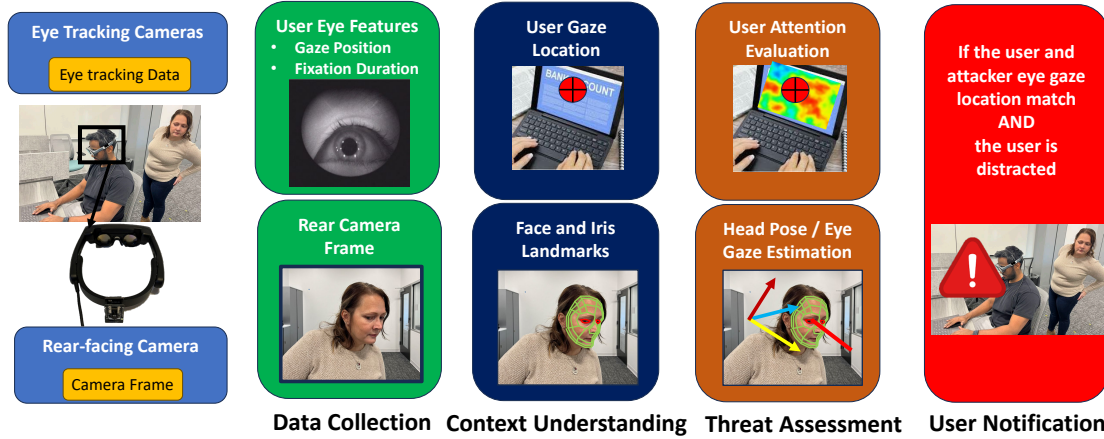


Fig. 3. SHOULДАР Design. The design, shown from left to right, begins by harnessing the AR device's onboard sensors and an additional rear-facing camera. After this, in Step 1 (Data Collection), the raw data is collected for both the user's and the potential attacker's position and eye gaze. This data is analyzed in Step 2 (Context Understanding) to determine the user's gaze location and the potential attacker's facial and eye landmarks. In Step 3 (Threat Assessment), this context is analyzed, and the user is alerted if they are deemed vulnerable to an SSA in Step 4 (User Notification).

metadata from the front-facing camera that gives us the user's current location and head pose (e.g., the roll, pitch, and yaw of the user's head) for each of the captured rear-facing frames. This information will be used later to locate the device and adjust rear-facing camera detections according to the position and rotation of the user's head. Finally, using the onboard eye-tracking cameras, we constantly estimate the user's eye gaze. We can also extract eye-tracking features (e.g., the fixation duration) identified as important to our design in Section 3.2.

Step 2: Context Understanding. In this step, we use the information collected from the headset and the rear-facing camera to develop context about the scenario. We can gather the user's position and head pose from the AR headset. Additionally, with the benefit of the AR headset's eye tracking, we know where the user is looking in real time. From the rear-facing camera frame, we use a facial detection inference model (e.g., MediaPipe [27] or OpenCV [54]) to locate any face in the vulnerable area. This inference model must have enough detail in its output to accurately calculate the *Perspective in Point (PnP)* of the face [65].

One output of the *PnP* calculation is the rotation of the face, needed to estimate the potential attacker's head pose. Using the localization and head pose of the user, combined with the rotation and local position of the face in the rear-facing camera, we can calculate the position and orientation of the face to the user's rear. We also require information about the person's iris position, pupil position, or other information indicating the direction of the eye [73]. Facial detection models (e.g., MediaPipe [27]) has such information with the output of the model.

Step 3: Threat Assessment. To reduce false positives, we harness the user's eye gaze in another way to determine when they are vulnerable to a threat. As mentioned in Section 3.2, fixation duration gives us insight into the concentration level of the user. Using the device's onboard eye-tracking sensors, we can measure the average fixation duration of the user and compare this against a threshold (discussed in more detail in Section 5). We can then estimate the direction of the attacker's gaze and notify the user of a threat when it aligns with the user's area of focus.

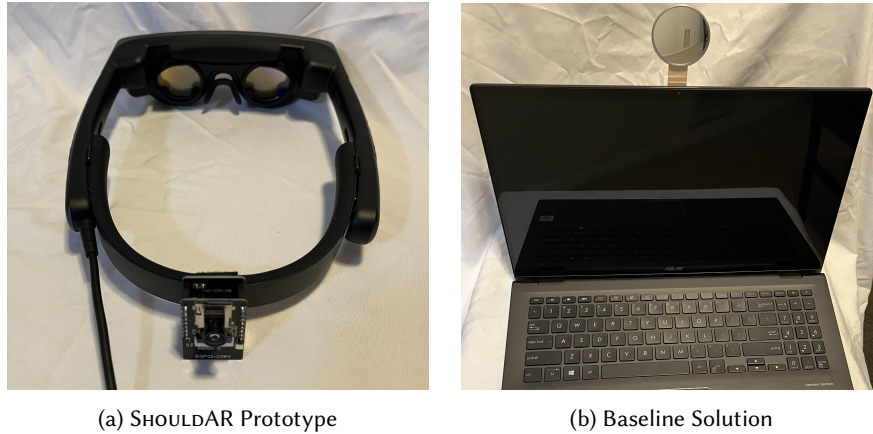


Fig. 4. (a) shows an image of the SHOULDAR prototype. A Magic Leap 2 headset was used as the AR platform, and the rear-facing camera, shown at the rear of the headset, was an ESP32-CAM and OV2640 wide-angle camera attached using Velcro fasteners. (b) provides an image of the baseline solution provided for participants. This two-inch fisheye convex mirror was used in both the typing and writing scenarios and was intended to simulate an ATM mirror. Participants were allowed to adjust these mirrors to see the entire testing area and could continue to adjust the mirror during the test if they wished.

Step 4: User Notification. Knowing the location of a potential attacker, their head pose and eye gaze, and the position and attention level of the user, we can combine this information to determine (1) if a person is positioned in a vulnerable area behind the user, (2) the target of the device user’s eye gaze and their attention level, and (3) if the person behind the user is looking over the shoulder of the user towards the same target. Applying minimum values to the distance between the user and the potential attacker’s eye gaze and the fixation duration of the user, we can also create gaze overlap and attention thresholds. If these thresholds are met (e.g., a person behind the user is looking over their shoulder while the user is potentially unaware), we alert the user to the threat.

5 Prototype Implementation

To understand the usability and effectiveness of the SHOULDAR design outlined in Section 4, we implemented the design using a Magic Leap 2 (ML2) AR headset [40]. This section outlines the technical specifics that were used to realize the SHOULDAR design presented in Section 4.

We developed the prototype using Unity version 2022.2.19f1. All code was written in C and C#. To implement the design outlines in Step 1 (Data Collection), we implemented the SHOULDAR design using an ML2 AR headset, as shown in Fig. 4a. This device has integrated eye-tracking sensors and also provides 3D spatial understanding through simultaneous localization and mapping (SLAM) [40]. The rear-facing camera requirement was satisfied using an ESP32-CAM microcontroller and an OV2640 camera with a wide-angle (140°) fish eye lens [8, 25]. A socket connection was created from the ESP32-CAM to the ML2 using a TP-Link AC1900 wireless router [69]. Raw frames were captured using the ESP32 and OV2640 and transmitted using this socket for analysis on the headset itself.

For Step 2 (Context Understanding), each rear-facing camera frame’s facial detection and face and iris landmarks were computed using MediaPipe [27]. Our prototype did not require offloading facial inference, as the ML2 proved to be fully capable of conducting this task on-device with negligible impact on the device’s expected frame rate and an average of 25 ms inference time.

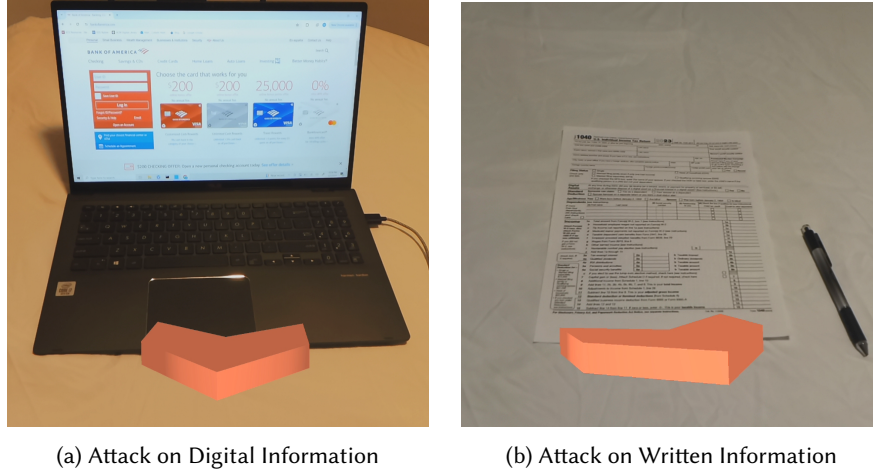


Fig. 5. SHOULDAR Notification Example. SHOULDAR uses a simple, red, 3D arrow to alert the user to a potential SSA and give the user the direction of the detected attack. In (a), an attack on a banking website login is occurring from the user’s rear. In (b), the user’s written tax information is being observed from their right rear.

In Step 3 (Threat Assessment), we used a 300-ms threshold for average user eye fixation duration as a minimum value to determine user attention. This threshold was decided during pilot testing in keeping with the maximum expected fixation duration when a human is not concentrating [44]. Above this, we expect “deeper cognitive processing” indicative of increased concentration. After determining the head pose (e.g., the rotation of the head), we use a simple method for determining the gaze direction of the potential attacker by estimating the position of the iris landmarks in relation to the eyelid landmarks. For example, if the attacker’s iris is positioned near the inside corner of the attacker’s left eye (e.g., the left iris is near the nose), we offset the estimation of the attacker’s head pose with 45° of gaze angle. This eye gaze estimation is done on the potential attacker’s face in each captured frame. We also require a time threshold to consider the overlap between the user and potential attacker’s gaze to alert the user that someone is observing their information. The 3D position of the user’s eye gaze is surrounded by an invisible, 3D sphere with a radius of 0.5 meters. If the attacker’s gaze intersects with the user’s gaze sphere for longer than the 300-ms threshold, the system provides an attack notification to the user.

Finally, in Step 4 (User Notification), we evaluate the user’s attention (using fixation duration) and the overlap of the user and potential attacker’s eye gaze locations. If conditions are met, we present the user with an arrow just below the center of their FOV. This arrow is oriented at the location of the detected attack and remains for three seconds, or as long as the attack is still being detected, whichever is longer. This notification is illustrated in Fig. 5. This red arrow was roughly 50% of the width of the user’s display, located inside the user’s FOV, and was designed not to be accidentally overlooked by the user while not completely obscuring their view of the sensitive material. The arrow is also a standard method to communicate direction, used here to indicate the direction of the SSA. While we do not test multiple notification techniques (which is not the main focus of this paper), we suggest possible alternatives in Section 7.

6 Evaluation

To understand the impact of the SHOULDAR prototype, we crafted an evaluation designed to measure the system’s effectiveness (through quantitative testing in various situations) and its usability. Before and after participants used the physical prototype, we asked them subjective questions about their comfort in different situations,

spanning multiple angles and distances inside the shoulder surfing vulnerable area. These questions were designed to understand the impact, if any, that active systems such as SHOULDAR might have on user comfort in potentially vulnerable situations. Additionally, we asked a series of questions designed to evaluate the usability of the design and prototype.

6.1 Evaluation Procedure

Our evaluation was designed to measure participant comfort with and without an active SSA protection system, and the effectiveness of the SHOULDAR system itself. This evaluation was approved by our university's Institutional Review Board (IRB).

Demographics. Participants were recruited from university graduate and undergraduate mailing lists and solicitations to local, non-student campus personnel and were compensated \$20 for approximately 45 minutes of time. Our evaluation included 24 participants across both the pilot (5 participants) and main study (19 participants). The pilot study results were used to determine system parameters such as fixation duration and were not included in the final evaluation. Of the 19 main study participants, 54.1% self-declared as male, while 45.8% self-declared as female. The ages of participants ranged from 19 to 39, with occupations ranging from nurses to graduate students at our university. Since the ML2 requires prescription corrective lenses specific to the device, participants did not wear glasses during the study. While 76.2% of participants were wearing neither contacts nor glasses during the testing, 23.8% were wearing contacts. The majority, 56.5%, of participants had never used AR before, or only once or twice.

Initial Participant Comfort Assessment. The evaluation was designed to explore the impact of an active, usable system on comfort in a situation vulnerable to an SSA. We began with introductions and signing an informed consent form. Then, we asked the participant to answer six questions about their comfort in six different scenarios at angles ranging from 0° to 40° and distance from 1 to 2 meters in .5 meter increments. These scenarios were shown as pictures and were demonstrated by a person physically standing at the angle and distance in the picture to give the participants a better sense of the different distances and angles. Participants were then also asked to imagine that they were either typing a password at a desktop computer (the *Typing* task) or writing sensitive information on a piece of paper at a desk (the *Writing* task) with a person at each of these angles and distances. Participants were asked to rate their comfort on a 5-point Likert scale ranging from 1 (Least Comfortable) to 5 (Most Comfortable) [45] with each of these situations.

SHOULDAR Prototype Evaluation. After completing the initial survey, participants were given a brief tutorial on using the ML2 and completed the ML2's fit and eye gaze calibration. After calibration, participants were asked to complete one of two tasks: data entry into an online typing assessment or completing a set of short answer questions using pen and paper. These tasks were designed to emulate the input of sensitive information during both typing and writing, similar to entering a password or writing an identification number or an address.

During each task, a simulated attacker moved behind the user to nine different positions in the evaluation area. The test environment with the locations of the nine positions is illustrated in Fig. 6. The attacker remained in each of these positions for exactly 10 seconds. Each of the nine positions included evenly-distributed variations in attacker distance, offset (e.g., angle from participant), head position, and eye gaze angle. At each distance (1m, 1.5m, and 2m), the attacker's head and eyes were directed away from the participant's information, head and eye directed at the participant's information, and with the attacker's head rotated away, but eyes directed toward the information. These tests were designed to test for false negatives, and the usefulness of head pose alone and with eye gaze to understand the added value of using the attacker's eye gaze. The nine positions of the attacker were repeated for both the typing and writing tasks across the three individual test types (i.e., baseline and two versions of SHOULDAR) for each participant in the same order, for a total of 54 attacker positions behind each

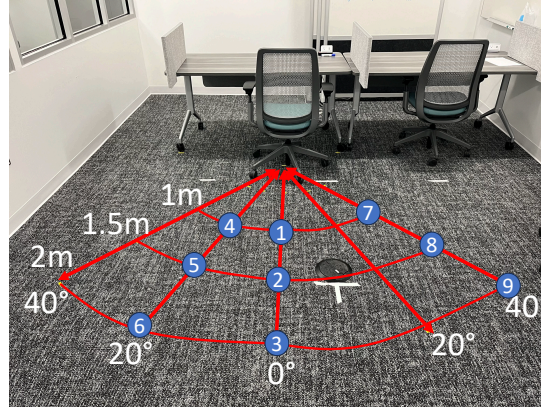


Fig. 6. An image of the testing environment. The red lines illustrate the distances and angles we evaluated inside the vulnerable area. The blue dots show an example of the ordering of the nine positions of the potential attacker.

participant. The participant was instructed to report when they believed this attacker was observing their actions by either raising a hand or speaking the word “attack”. The participant was instructed to conduct their task (typing or writing) and report potential attacks simultaneously. Because the ML2 required an eye gaze calibration, every participant was asked to evaluate the baseline solution first, and then the head pose and eye gaze enabled SHOULДАР prototypes. This was to reduce the requirement for recalibration of the device if removed between tests.

The baseline required the user to determine the presence of an attack using their understanding of the area to their rear. During the two SHOULДАР prototype tests, participants were asked to report when the prototype indicated an attack on their display. This prevents any learning effect from the participant watching the attacker in the baseline. For the baseline, the participant was provided with a small (two-inch diameter) fisheye convex mirror attached to the laptop at their desk (see Fig 4b for an illustration). This mirror was designed to emulate an ATM mirror discussed in Section 2.2, a widely used solution to the SSA problem that could be used in both typing and writing scenarios [41, 51]. Other solutions are designed for only a single situation, like mobile devices [10]. The participants were allowed to reposition this baseline solution as they saw fit during this part of the evaluation to observe the entire testing area behind them. They were also allowed to adjust the mirror as they saw fit during the test to adjust for their movement or repositioning. For each test, the attacker moved to 18 positions for both the typing and writing tests (9 for each). There were 54 tests for the baseline and two versions of SHOULДАР.

Final Participant Assessment. The participants were asked to complete a final questionnaire following the prototype evaluation. The first six scenario descriptions were nearly identical to those presented in the initial survey but asked the participant to imagine these scenarios *while wearing the prototype they had just evaluated*. This was designed to gauge any change in the participant’s concern in these scenarios if protected by a system such as SHOULДАР. Participants were asked eight additional questions designed to assess the prototype’s usability (see Table 3). Participants were asked to rate their agreement or disagreement with each question using a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

We provide an evaluation of SHOULДАР in three parts. First, Section 6.2 shows the results of the Likert responses of the participants before and after experiencing a mobile, usable protection system such as SHOULДАР. Second, Section 6.3 gives quantitative results on the effectiveness of SHOULДАР, expressed as true positive and false negative rates in various scenarios. Finally, we seek to understand how participants felt about the prototype’s

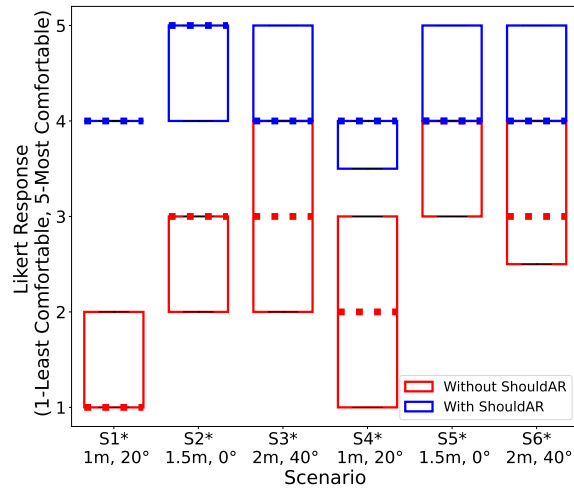


Fig. 7. A comparison of the Likert responses of participants when asked about their comfort in a variety of SSA-vulnerable situations. The box plots in red are the questions asked assuming no protection. The blue box plots are the questions asked assuming protection from an active system such as SHOULDAR. The dotted lines show the median values. Scenarios with an asterisk (*) indicate statistically significant results ($p < 0.05$). The x-axis is labeled with the attack distances and angles of each scenario, respectively.

usability. Section 6.4 outlines the qualitative responses from users when asked to agree or disagree with eight usability statements.

6.2 Understanding SHOULDAR's Impact on Comfort in SSA-vulnerable Situations

Before and after evaluating the SHOULDAR prototype, participants were asked to rate their comfort in six scenarios on a 5-point Likert scale. The scenarios consisted of three different angles at three different distances while the user completed one of two tasks (typing or writing). Fig. 7 shows the responses for the six scenarios before (in red) and after testing the prototype (in blue). Overall, *participants were more comfortable in each situation while wearing an active protection system, such as SHOULDAR, than they were without*. Intuitively, participants were more concerned with the attack scenarios where the attacker is close (S1 and S4) than the longer-range scenarios. Generally, SHOULDAR increased median responses for participant comfort from neutral or lower to positive responses, with the largest individual scenario changes in the close-range and typing scenarios. For each question, a Wilcoxon signed rank test showed statistical differences in the pre-evaluation and post-evaluation Likert results ($p < 0.05$) [70].

Fig. 8 breaks down the participant comfort analysis by self-declared gender. We can see that participants, regardless of gender, reported an increase in comfort while protected by an active system such as SHOULDAR. For male participants, SHOULDAR increased median responses for participant comfort from neutral or lower to positive responses; all questions showed statistical significance ($p < 0.05$), except Q5 ($p = 0.18$). Female participants also increased median responses for participant comfort from neutral or lower to positive responses considering a prototype like SHOULDAR; all questions showed statistical significance ($p < 0.05$).

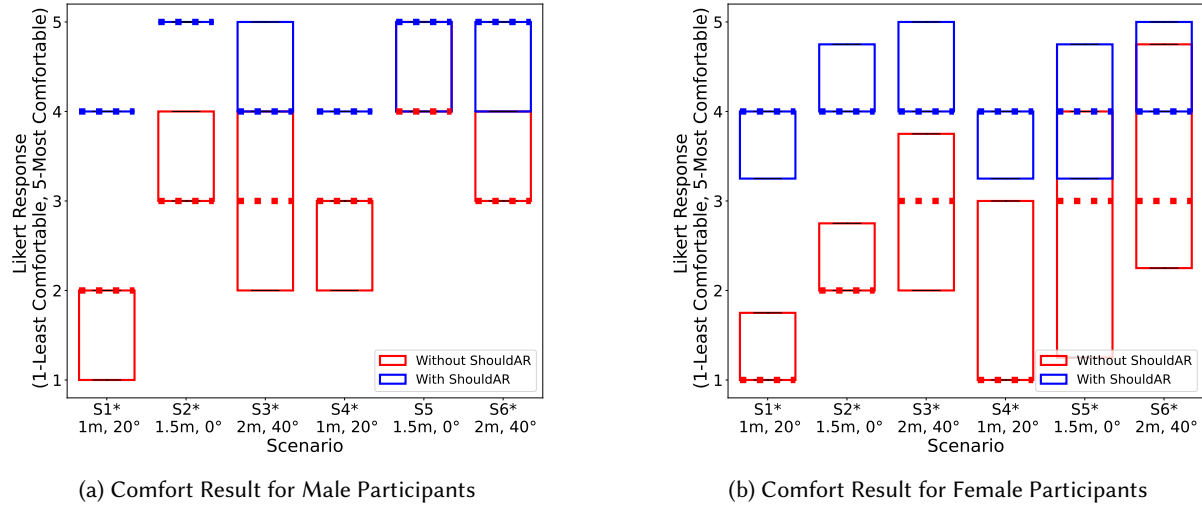


Fig. 8. A comparison of the Likert responses of participants when asked about their comfort in the situations outlined in the initial comfort survey, separated by participant gender. (a) is the responses from self-declared male participants. (b) is from self-declared females. The box plots in red are the questions asked assuming no protection. The blue box plots are the questions asked assuming protection from an active system such as SHOULДАР. The dotted lines show the median values. Scenarios with an asterisk (*) indicate statistically significant results ($p < 0.05$).

User Actions	Baseline		Head Pose (HP)		HP w/ Eye Gaze	
	Typing	Writing	Typing	Writing	Typing	Writing
True Positive	43.86%	41.23%	58.77%	53.51%	89.47%	85.09%
Total True Positive	42.54%		56.14%		87.28%	
False Positive	0.88%	0.88%	0.00%	1.75%	0.88%	0.00%
Total False Positive	0.88%		0.88%		0.44%	

Table 2. A summary of the results of the SHOULДАР quantitative evaluation. The “Baseline” column represents the tests conducted using the two-inch diameter fisheye mirror. The “Head Pose (HP)” column shows the tests using only the attacker’s head pose and not eye gaze. “HP w/ Eye Gaze” shows the results using both head pose and eye gaze. Each modality is shown subdivided by individual test tasks (e.g., typing or writing) and as a total average.

6.3 Evaluating SHOULДАР’s Effectiveness

Next, we seek to evaluate SHOULДАР quantitatively by analyzing the True Positive (TP) and False Positive (FP) rates recorded during the evaluation. Section 5 outlines the participant tasks and attacker positions evaluated. The baseline solution consisted of a two-inch diameter fisheye mirror positioned by the participant during both the typing and writing tasks. This baseline is compared to two versions of the SHOULДАР prototype, one using both attacker head pose and eye gaze estimations and one only using head pose. This was designed to understand the effect of the attacker’s eye gaze on the results.

Table 2 provides a summary of the results of these three tests. The baseline (mirror) solution showed the lowest TP rate with an average of 42.54% of attacks being detected by the participant. The FP rate of the baseline was also nearly one percent, equal to the head-pose-enabled solution. The majority of the False Negative (FN) results,

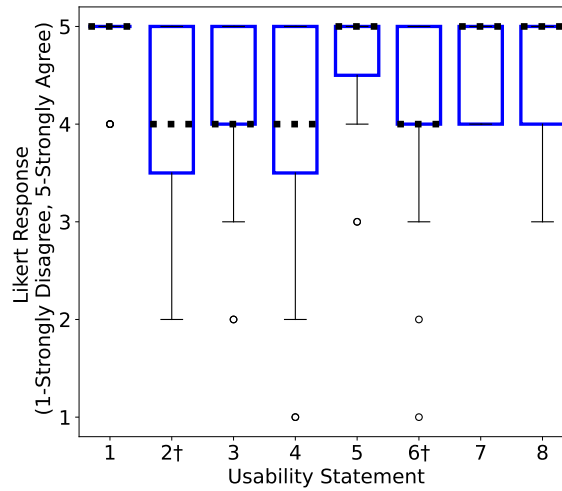


Fig. 9. A comparison of the Likert responses of participants when asked about their comfort in the situations outlined in Table 3. The dotted lines show the median values. **Note:** Questions 2 and 6 (marked with a †) were worded in such a way as to ensure the participant was carefully reading the questions. The Likert values for these questions were inverted in this plot for easy comparison with the remaining questions.

where the participant failed to indicate an ongoing attack, occurred when the participant obscured the view from the mirror.

The head-pose-only solution, using only the direction of the potential attacker’s head to decide if they are attempting an SSA, improved on the baseline performance with a TP average of 56.14%. The majority of the FN results for the head pose method occurred when the attacker’s head was directed *away* from the participant’s task, but their eyes were directed *toward* the task. This shows the vulnerability of such a solution to these scenarios.

The eye gaze solution, including both the attacker’s head pose and eye gaze to detect an attack, performed best. This version of the prototype averaged an 87.28% TP rate over both writing and typing tasks. While not perfect, this shows roughly a *two-fold increase in TP rate over the baseline, and a 31% increase over the head-pose-only solution*. This 31% increase in TP rate is directly related to the successful head-away/eyes-toward scenarios that hampered the head pose-only version. The majority of the FN tests (the tests where the attacker was observing the participant’s information but the participant did not indicate an attack) occurred at the 2 meter attack distance (18 out of 29 total false negatives). This shows a limitation of this technique, as small errors in eye gaze estimation are magnified at larger distances from the participant. Previous work shows, however, that distances farther than 2 meters are less vulnerable to SSAs [10], making this less of a concern. More accurate eye gaze estimation methods can also improve this system, as discussed in Section 7. Additionally, the FP rate of this system was less than 0.5%, resulting in only a few unnecessary warnings to the participant.

6.4 Evaluating SHOULDAR’s Usability

Finally, we evaluate the results of the usability-focused questions posed to the participants. Fig. 9 shows the results of the questions outlined in Table 3. Questions 1, 3, 4, 5, 7, and 8 were all designed to elicit a higher Likert score if the user believed the prototype was usable in general. In contrast, Q2 and Q6 were designed on an opposite scale to ensure participants read questions carefully. Some questions were straightforward. Question 1 asked participants if they believed the prototype was easy to use. Question 2 asked participants if they believed

Table 3. Participants were asked to rate their agreement with the following statements or each question. A 5-point Likert scale was given, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Question	Usability Evaluation Statements
1	The prototype you evaluated was easy to use.
2	The prototype was overly intrusive on your activities.
3	The requirements of the prototype (e.g., wearing the headset) were worth the benefit I received.
4	If made smaller and more lightweight, I would wear this system as part of my daily routine.
5	I prefer the prototype’s level of protection to methods such as the mirror to inform me of shoulder surfing attacks.
6	I believe that using a method to defend against Shoulder Surfing Attacks that require direct observation (such as a mirror) is more usable than the prototype.
7	I would use this system if made less cumbersome and integrated into future AR devices.
8	I felt more secure against shoulder surfing attacks using the prototype than I would with any other method I know of.

the prototype was “overly intrusive”. We allowed participants to decide what “overly intrusive” meant to them. Other questions were more speculative and should be given less weight in determining usability. Questions 4 and 7 assumed a future prototype designed without a Velcro-mounted rear-facing camera and was lighter in weight, keeping with our future assumptions in Section 1. Since such systems do not yet exist, we leave these questions open for interpretation.

In conversations after the evaluation, users expressed that the prototype allowed them to concentrate on the writing or typing task far more easily than the baseline (mirror) solution. The baseline required them to evaluate their vulnerable area constantly and distracted them from their task. Knowing that SHOULДАР would alert them in the case of a potential attack, participants expressed a general preference over the baseline, shown empirically in the usability evaluation results.

7 Discussion

To address a lack of usable solutions to SSAs, we designed, implemented, and evaluated SHOULДАР. This system harnesses the eye gaze information from both the potential attacker and victim to determine if there is likely an SSA and to give the victim alert information that can be used to stop it. Our evaluation determined that such a system improved the perception of comfort when participants were asked to imagine potentially vulnerable situations. Additionally, we proved that such a system can accurately detect SSAs with a low FP rate (e.g., less than 1%) in the scenarios we evaluated.

Potential Privacy Concerns. SHOULДАР uses open-source *facial detection* techniques to identify faces to the user’s rear. Additional information from these detections is used to analyze these faces’ head pose and eye gaze. SHOULДАР does not retain this information, as seen in *facial recognition* technologies that attempt to use this information to re-identify faces later in time [20]. Certain large-scale facial recognition techniques have already been deemed impermissible under the European General Data Protection Regulation (GDPR) and the EU AI Act [1]. They would make a system relying on these techniques difficult without concern for legal consequences [18]. While SHOULДАР uses a less-intrusive method of identifying faces, questions remain about whether techniques like SHOULДАР are acceptable in places with large crowds (e.g., airports) and whether these techniques violate local or national law. For example, previous work has shown that similar facial detection in public billboards may or may not violate the GDPR [20].

For this reason, we seek to highlight likely use cases for SHOULDAR. Our system is likely more permissible in sensitive areas where privacy expectations for the general public are already reduced (e.g., in banks, government offices, and ATMs). In these sensitive areas and situations, users and those around the device already understand the possibility that their images and faces may be used in certain circumstances. In other situations, such as in cafes or airports, systems like SHOULDAR may infringe on the privacy of those in the rear-facing cameras collection area (e.g., the user's vulnerable area). Additionally, any system similar to SHOULDAR should take great care not to move from facial detection to recognition, as this has been shown to be legally impermissible in certain contexts [18].

What if the Attacker's Eyes Are Not Visible? SHOULDAR matches the user's eye gaze with the potential attacker's eye gaze in 3D. This eye gaze is necessary to ensure a fine-grained understanding of the focus of the user and attacker's attention. Without it, the system would likely suffer from increased false positives as isolating data pertinent to the user would be difficult. However, the inference model analyzing the rear-facing camera frames may detect a face but not the person's eyes. This can be due to intentional or unintentional obscuration with the person's hand, sunglasses, AR device, or other means. If the prototype detects a face but does not have the benefit of the eye-gaze estimation, the prototype defaults to head-pose-only estimation. Research into the correlation of head pose and eye gaze shows a strong relationship [74], but not so strong as to be able to use head pose alone to isolate the user's attention. Eye shift alone (i.e., without head movement) can be expected when the shift in focus is less than 20° to 30° [66]. This shows that while head movement can estimate the user's attention, their eyes give the exact location with less error and less ambiguity. However, using the head pose only as a last resort is sufficient when the attacker's eyes are in the same direction as the head, but it results in a reduced TP rate when the head and eyes are not aligned. Last, if the attacker intentionally wears a mask to obscure their face from facial detection and head pose estimation, more advanced sensors such as thermal cameras could potentially be used to provide an alternative method of estimating head pose [34].

SSA Threats from Outside the Vulnerable Area. Our vulnerable area is designed to protect from attacks on mobile devices, laptops/desktops, and written documents, using feasible distances from earlier work [10]. However, if the user is typing or writing sensitive information on a whiteboard or large display, SSAs could be feasible from greater distances. While Section 6 showed less comfort when the SSA attacker was positioned closer, a system that can assess from a greater distance would remain desirable.

As mentioned in Section 4, we acknowledge that the vulnerable area could be larger when the victim concentrates on an area directly at the center of their FOV. This would limit the expectation that they would be using the full mobility of their eyes to their left and right to scan their visual field for attackers. Another limitation comes from detecting attacks on large displays where the sensitive information is visible to the attacker without intersecting with the user's gaze. However, we expect active SSA systems to be used in sensitive environments where large displays are less likely. We imagine scaling up the number of cameras attached to the headset to expand the SSA detection range in future designs. Finally, ShouldAR is not designed to detect SSAs from the user's front, as it is far more likely the device user would notice these attacks, and more likely that the content could be occluded by a display (i.e., a monitor).

Calculating the Attacker's Eye Gaze. There are many ways to estimate a person's eye gaze from an image of their face. Techniques such as [42, 55] use the neural networks, trained on labeled images of specific eye orientations, to determine the gaze angle. Our prototype sought a lighter, faster method that did not involve neural networks. Using a rough estimation of the gaze angle based on the distance of the iris from the eyelid landmarks, we could estimate eye gaze using hard-coded methods such as [19]. This is not as accurate as using neural networks (e.g., a convolutional neural network), but we believe it to be less likely to suffer from delay due to the inference time required for each frame. Other methods that could create a more fine-grained estimation of

the eye angle, such as [64], also exist. These were not used as part of our prototype. Our test used a maximum attacker distance of 2 meters for the reasons described in Section 3.2. At farther ranges, this course-grained estimation method may begin to fail. This limitation is less of a concern in light of previous work explaining the limited feasibility of SSAs from beyond 2 meters [10].

Attack Notification Techniques. Our evaluation was designed to measure the user comfort resulting from using our prototype and the system’s effectiveness at detecting SSAs. While wearing an ML2 during the evaluation, future studies into designing optimal notification cues for SSAs and evaluating user perceptions are necessary. Additionally, as mentioned in Section 1, there are many other devices that can potentially provide some or all of the sensors required for our design (e.g., Meta’s Ray-Ban and Project Aria glasses [47, 48]). With modifications, devices without holographic displays could use haptics or audio feedback to notify users, and devices without user eye gaze sensors could simply alert the user if someone was looking in their general area. While these devices are feasible, we still believe that AR headsets are currently the most viable and available solution.

Future Work. In future work, we plan to explore improving system accuracy at or beyond 2 meters to address SSAs from greater distances. Additionally, we plan to develop user studies to investigate an optimal method for alerting the device user once an SSA is detected. This work could analyze other visualizations or multimodal (e.g., audio or haptics) alerts. Little research effort was given to determining and analyzing the best methods to alert the device user to an SSA once one is detected. The simple 3D arrow used in our prototype was not the focus of rigorous analysis or investigation. Our future work will analyze other, possibly less intrusive or overt methods to alert the user to an SSA.

Additionally, the device user cannot declare that they are aware of and consent to a specific person’s observation of their sensitive data. Currently, the user must halt SHOULДАР temporarily if they wish to prevent false positives in this scenario. There are potentially more intuitive ways to give a user the ability to label a face as “trusted” and to ignore potential attacks from this person. Specifically, creating a facial recognition system allows the user to select certain faces and compare these “trusted” faces to each facial detection. This detection could be ignored if the faces match (e.g., the face is a “friend” or a “collaborator” on a project with a common display).

8 Conclusion

To address a lack of usable solutions to the SSA problem, we designed, implemented, and evaluated SHOULДАР, a mobile, usable anti-SSA system. This system uses an AR headset and multimodal eye gaze information from both the headset user and the potential attacker. SHOULДАР is usable in a wide variety of scenarios, including when writing, using an ATM PIN pad, or using a mobile device. This system was shown to be preferable over a widely used anti-SSA solution, a rear-facing mirror. While AR headsets are not as ubiquitous as mobile phones yet, we believe that such a solution would be feasible. After improvements in weight and battery life, a solution such as SHOULДАР would be effective at reducing user concern in SSA-vulnerable scenarios.

Acknowledgments

This work is supported in part by the Commonwealth Cyber Initiative (CCI) and an NSF Grant under CNS 2153397.

References

- [1] 2024. Artificial Intelligence Act: MEPs adopt landmark law. Retrieved April 15, 2024 from <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>
- [2] Ahmed Abdou, Ekta Sood, Philipp Müller, and Andreas Bulling. 2022. Gaze-Enhanced Crossmodal Embeddings for Emotion Recognition. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 138 (may 2022), 18 pages. <https://doi.org/10.1145/3530879>

- [3] Yasmeen Abdrabou, Mohamed Khamis, Rana Mohamed Eisa, Sherif Ismail, and Amr Elmougy. 2019. Just Gaze and Wave: Exploring the Use of Gaze and Gestures for Shoulder-Surfing Resilient Authentication. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) (ETRA '19). Association for Computing Machinery, New York, NY, USA, Article 29, 10 pages. <https://doi.org/10.1145/3314111.3319837>
- [4] Mohammed Eunus Ali, Anika Anwar, Ishrat Ahmed, Tanzima Hashem, Lars Kulik, and Egemen Tanin. 2014. Protecting Mobile Users from Visual Privacy Attacks. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) (UbiComp '14 Adjunct). Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/2638728.2638788>
- [5] Mohammed Eunus Ali, Anika Anwar, Ishrat Ahmed, Tanzima Hashem, Lars Kulik, and Egemen Tanin. 2014. Protecting Mobile Users from Visual Privacy Attacks. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) (UbiComp '14 Adjunct). Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/2638728.2638788>
- [6] Maraim Alnefaie and Srinivas Sampalli. 2019. BraillePassword: accessible web authentication technique on touchscreen devices. *Journal of Ambient Intelligence and Humanized Computing* 10 (06 2019). <https://doi.org/10.1007/s12652-018-0860-x>
- [7] Suliman A. Alsuhibany. 2022. Vibration-Based Pattern Password Approach for Visually Impaired People. *Comput. Syst. Sci. Eng.* 40, 1 (2022), 341–356. <https://doi.org/10.32604/csse.2022.018563>
- [8] Amazon. 2023. Treedix OV2640 Camera Module 140 Degree Wide Angle CMOS 2MP Camera Mini Camera Module. Retrieved Dec 16, 2023 from https://www.amazon.com/Treedix-OV2640-Camera-Module-Degree/dp/B0894KKXHX/ref=asc_df_B0894KKXHX/?tag=hyprod-20&linkCode=df0&hvadid=658532215676&hvpos=&hvnetw=g&hvrnd=13097930281404645000&hvpone=&hvpstwo=&hqvmt=&hqvdev=c&hvdvcmdl=&hvlolocin=&hvlolocphy=9008695&hvtargid=pla-1489482268684&psc=1&mcid=6741e1d20ca03b7ba63f7457dfe91ada
- [9] Apple. 2023. *Apple Vision Pro*. Apple Corporation. Retrieved Dec 11, 2023 from <https://www.apple.com/apple-vision-pro/>
- [10] M. Bace, A. Saad, M. Khamis, S. Schneegass, and A. Bulling. 2022. PrivacyScout: Assessing Vulnerability to Shoulder Surfing on Mobile Devices. In *Privacy Enhancing Technologies Symposium 2022 (PETS 2022)* (Sydney, Australia). <https://doi.org/10.56553/popets-2022-0090>
- [11] Divya Bhati. 2023. ATM card scams: What is Shoulder Surfing and how to protect your ATM PIN from scammers. India Today. Retrieved Dec 11, 2023 from <https://www.indiatoday.in/technology/news/story/atm-card-scams-what-is-shoulder-surfing-and-how-to-protect-your-atm-pin-from-scammers-2402309-2023-07-05>
- [12] Andrea Bianchi, Ian Oakley, Vassilis Kostakos, and Dong Soo Kwon. 2010. The phone lock: audio and haptic shoulder-surfing resistant PIN entry methods for mobile devices. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*. 197–200.
- [13] Robert Biddle, Sonia Chiasson, and P.C. Van Oorschot. 2012. Graphical Passwords: Learning from the First Twelve Years. *ACM Comput. Surv.* 44, 4, Article 19 (sep 2012), 41 pages. <https://doi.org/10.1145/2333112.2333114>
- [14] Basak Bilgi and Bulent Tugrul. 2018. A Shoulder-Surfing Resistant Graphical Authentication Method. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*. 1–4. <https://doi.org/10.1109/IDAP.2018.8620934>
- [15] Andrew Lim Chee Yeung, Bryan Lee Weng Wai, Cheng Hao Fung, Fiza Mughal, and Vahab Iranmanesh. 2015. Graphical password: Shoulder-surfing resistant using falsification. In *2015 9th Malaysian Software Engineering Conference (MySEC)*. 145–148. <https://doi.org/10.1109/MySEC.2015.7475211>
- [16] Chun-Yu (Daniel) Chen, Bo-Yao Lin, Junding Wang, and Kang G. Shin. 2019. Keep Others from Peeking at Your Mobile Device Screen!. In *The 25th Annual International Conference on Mobile Computing and Networking* (Los Cabos, Mexico) (MobiCom '19). Association for Computing Machinery, New York, NY, USA, Article 22, 16 pages. <https://doi.org/10.1145/3300061.3300119>
- [17] Shudi Chen and Youwen Zhu. 2019. A Textual Password Entry Method Resistant to Human Shoulder-Surfing Attack. In *Cyberspace Safety and Security*, Jaideep Vaidya, Xiao Zhang, and Jin Li (Eds.). Springer International Publishing, Cham, 409–420.
- [18] CNIL. 2019. Facial recognition: for a debate living up to the challenges. Retrieved April 9, 2024 from <https://www.cnil.fr/en/facial-recognition-debate-living-challenges#:~:text=Facial%20recognition%20is%20raising%20new,and%20global%20public%20agendas%20alike.>
- [19] Asadullah Dal. 2023. Eyes-Position-Estimator-Mediapipe. Retrieved Feb 2, 2024 from <https://github.com/Asadullah-Dal17/Eyes-Position-Estimator-Mediapipe>
- [20] P. A. Earls Davis. 2020. Facial Detection and Smart Billboards: Analysing the 'Identified' Criterion of Personal Data in the GDPR. *European Data Protection Law Review* (2020). <https://api.semanticscholar.org/CorpusID:226642930>
- [21] Alexander De Luca, Katja Hertzschuch, and Heinrich Hussmann. 2010. ColorPIN: Securing PIN entry through indirect input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1103–1106.
- [22] Alexander De Luca, Roman Weiss, and Heiko Drewes. 2007. Evaluation of Eye-Gaze Interaction Methods for Security Enhanced PIN-Entry. In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces* (Adelaide, Australia) (OZCHI '07). Association for Computing Machinery, New York, NY, USA, 199–202. <https://doi.org/10.1145/1324892.1324932>

- [23] Reyhan Düzgün, Peter Mayer, and Melanie Volkamer. 2022. Shoulder-Surfing Resistant Authentication for Augmented Reality. In *Nordic Human-Computer Interaction Conference (Aarhus, Denmark) (NordiCHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 29, 13 pages. <https://doi.org/10.1145/3546155.3546663>
- [24] Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (*CHI '17*). Association for Computing Machinery, New York, NY, USA, 4254–4265. <https://doi.org/10.1145/3025453.3025636>
- [25] Espressif. 2023. Espressif - ESP32. Retrieved Dec 16, 2023 from <https://www.espressif.com/en/products/socs/esp32>
- [26] Dylan Gilbert and Michael Fagan. 2024. Journey into the Immersive Frontier: Preliminary NIST Research on Cybersecurity and Privacy Standards for Immersive Technologies. Retrieved Feb 13, 2024 from <https://www.nist.gov/blogs/cybersecurity-insights/journey-immersive-frontier-preliminary-nist-research-cybersecurity-and>
- [27] Google. 2023. MediaPipe. Retrieved Dec 16, 2023 from <https://developers.google.com/mediapipe>
- [28] Google. 2024. Google Scholar Search. Retrieved Feb 13, 2024 from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C47&as_vis=1&q=%22shoulder+surfing%22&btnG=
- [29] Michael Gorthaus. 2024. Vision Pro sales are really tanking, new supply chain data shows. Retrieved Apr 29, 2024 from <https://www.fastcompany.com/91112727/vision-pro-sales-tanking-supply-chain-data-meta-quest>
- [30] Ashtha Goyal, Priya Matta, and Yogesh Lohumi. 2024. Preventing Shoulder Surfing Attacks Matrix Based Graphical Technique. 53–58. <https://doi.org/10.1109/ICWITE59797.2024.10502656>
- [31] I.P. Howard and B.J. Rogers. 1995. *Binocular Vision and Stereopsis*. Oxford University Press. <https://books.google.com/books?id=I8vqITdETe0C>
- [32] Mohamed Khamis, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. 2016. GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI EA '16*). Association for Computing Machinery, New York, NY, USA, 2156–2164. <https://doi.org/10.1145/2851581.2892314>
- [33] Mohamed Khamis, Mariam Hassib, Emanuel von Zezschwitz, Andreas Bulling, and Florian Alt. 2017. GazeTouchPIN: Protecting Sensitive Data on Mobile Devices Using Secure Multimodal Authentication. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) (*ICMI '17*). Association for Computing Machinery, New York, NY, USA, 446–450. <https://doi.org/10.1145/3136755.3136809>
- [34] Marcin Kopaczka, Justus Schock, Jan Nestler, Kevin Kielholz, and Dorit Merhof. 2018. A combined modular system for face detection, head pose estimation, face tracking and emotion recognition in thermal infrared images. In *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 1–6.
- [35] Katharina Krombholz, Heidelinde Hobel, Markus Huber, and Edgar Weippl. 2015. Advanced social engineering attacks. *Journal of Information Security and Applications* 22 (2015), 113–122. <https://doi.org/10.1016/j.jisa.2014.09.005> Special Issue on Security of Information and Networks.
- [36] Romina Kühn, Mandy Korzetz, and Thomas Schlegel. 2019. User Strategies for Mobile Device-Based Interactions to Prevent Shoulder Surfing. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia* (Pisa, Italy) (*MUM '19*). Association for Computing Machinery, New York, NY, USA, Article 43, 5 pages. <https://doi.org/10.1145/3365610.3368412>
- [37] Manu Kumar, Tal Garfinkel, Dan Boneh, and Terry Winograd. 2007. Reducing Shoulder-Surfing by Using Gaze-Based Password Entry. In *Proceedings of the 3rd Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania, USA) (*SOUPS '07*). Association for Computing Machinery, New York, NY, USA, 13–19. <https://doi.org/10.1145/1280680.1280683>
- [38] Taekyoung Kwon and Jin Hong. 2014. Analysis and improvement of a pin-entry method resilient to shoulder-surfing and recording attacks. *Ieee transactions on information forensics and security* 10, 2 (2014), 278–292.
- [39] Otto Lappi. 2016. Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neuroscience & Biobehavioral Reviews* 69 (2016), 49–68.
- [40] Magic Leap. 2023. *magic Leap 2 - Device*. Magic Leap. Retrieved Dec 11, 2023 from <https://www.magicleap.com/magic-leap-2>
- [41] The Florida Legislature. 2023. The 2023 Florida Statutes (including Special Session C). *OJ Title XXXVII* (2023). Retrieved Dec 11, 2023 from http://www.leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=0600-0699/0655/Sections/0655.962.html
- [42] Oliver Lorenz and Ulrike Thomas. 2019. Real Time Eye Gaze Tracking System using CNN-based Facial Features for Human Attention Measurement. 598–606. <https://doi.org/10.5220/0007565300002108>
- [43] Feiyu Lu, Leonardo Pavanatto, and Doug A. Bowman. 2023. In-the-Wild Experiences with an Interactive Glanceable AR System for Everyday Use. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction* (, Sydney, NSW, Australia,) (*SUI '23*). Association for Computing Machinery, New York, NY, USA, Article 11, 9 pages. <https://doi.org/10.1145/3607822.3614515>
- [44] Bhanuka Mahanama, Yasith Jayawardana, Sundararaman Rengarajan, Gavindya Jayawardena, Leanne Chukoskie, Joseph Snider, and Sampath Jayarathna. 2022. Eye Movement and Pupil Measures: A Review. *Frontiers in Computer Science* 3 (2022). <https://doi.org/10.3389/fcsc.2022.895121>

- [//doi.org/10.3389/fcomp.2021.733531](https://doi.org/10.3389/fcomp.2021.733531)
- [45] Saul McLeod. 2019. Likert Scale Definition, Examples and Analysis. Retrieved Feb 2, 2024 from <https://www.simplypsychology.org/likert-scale.html>
 - [46] Tushar Mehta. 2023. *How AR glasses are going from niche gadget to smartphone replacement*. Digital Trends. Retrieved Dec 11, 2023 from <https://www.digitaltrends.com/mobile/ar-glasses-replace-smartphones-future-how/>
 - [47] Meta. 2024. Introducing Project Aria. Retrieved Apr 29, 2024 from <https://www.projectaria.com/>
 - [48] Meta. 2024. Shop Ray-Ban Meta Smart Glasses. Retrieved Apr 29, 2024 from <https://www.meta.com/smart-glasses/shop-all>
 - [49] Microsoft. 2024. Create images from words with AI. Retrieved Feb 13, 2024 from <https://www.bing.com/images/create/a-person-working-at-a-desk-on-a-laptop-with-a-person/1-65cb93bd21814fdd927fa22c3e8d0aa7?id=EowoJwMdjsIQj%2fjndAM3uw%3d%3d&view=detailedv2&idpp=genimg&idpbck=1&form=BICREC&idpview=singleimage&thid=OIG4.VXWiwzf4YiCerq6DxWjb>
 - [50] Shivsevak Negi and Ritayan Mitra. 2020. Fixation duration and the learning process: an eye tracking study with subtitled videos. *Journal of Eye Movement Research* 13 (08 2020). <https://doi.org/10.16910/jemr.13.6.1>
 - [51] Syeda Noor. 2024. Why Do ATMS Have Convex Mirror? Retrieved Feb 13, 2024 from <https://www.boldskey.com/insync/pulse/2016/why-do-atms-have-a-convex-mirror/articlecontent-pf137288-107799.html>
 - [52] J. O'Dwyer, N. Murray, and R. Flynn. 2019. Eye-based Continuous Affect Prediction. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE Computer Society, Los Alamitos, CA, USA, 137–143. <https://doi.org/10.1109/ACII.2019.8925470>
 - [53] Last 365 Days of SSA Research. 2019. Google Scholar Search. Retrieved Feb 2, 2024 from https://scholar.google.com/scholar?start=0&q=%22shoulder+surfing%22&hl=en&scisbd=1&as_sdt=0,47&as_vis=1
 - [54] OpenCV.org. 2023. MediaPipe. Retrieved Dec 16, 2023 from <https://opencv.org/>
 - [55] OpenVINO. 2023. gaze-estimation-adas-0002. Retrieved Feb 2, 2024 from https://docs.openvino.ai/2022.3/omz_models_model_gaze_estimation_adas_0002.html
 - [56] Romina Palermo and Gillian Rhodes. 2007. Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia* 45, 1 (2007), 75–92.
 - [57] Athanasios Papadopoulos, Toan Nguyen, Emre Durmus, and Nasir Memon. 2017. IllusionPIN: Shoulder-Surfing Resistant Authentication Using Hybrid Images. *IEEE Transactions on Information Forensics and Security* 12, 12 (2017), 2875–2889. <https://doi.org/10.1109/TIFS.2017.2725199>
 - [58] Alfredo J. Perez, Sherali Zeadally, Scott Griffith, Luis Y. Matos Garcia, and Jaouad A. Mouloud. 2020. A User Study of a Wearable System to Enhance Bystanders' Facial Privacy. *IoT* 1, 2 (2020), 198–217. <https://doi.org/10.3390/iot1020013>
 - [59] PonemonInstitute. 2016. Global Visual Hacking Experimental Study: Analysis. (2016). Retrieved Dec 11, 2023 from multimedia.3m.com/mws/media/12542320/global-visual-hacking-experiment-study-summary.pdf
 - [60] Reydar. 2023. The Future of Augmented Reality. Retrieved Dec 16, 2023 from <https://www.reydar.com/exploring-the-future-potential-of-augmented-reality-trends-technology-and-impact/#:~:text=In%20the%20near%20future%2C%20we,and%20technology%20than%20today's%20smartphones>
 - [61] Volker Roth, Kai Richter, and Rene Freidinger. 2004. A PIN-entry method resilient against shoulder surfing. In *Proceedings of the 11th ACM conference on Computer and communications security*. 236–245.
 - [62] Alia Saad, Dina Hisham Elkafrawy, Slim Abdennadher, and Stefan Schneeeggass. 2020. Are They Actually Looking? Identifying Smartphones Shoulder Surfing Through Gaze Estimation. In *ACM Symposium on Eye Tracking Research and Applications (Stuttgart, Germany) (ETRA '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 42, 3 pages. <https://doi.org/10.1145/3379157.3391422>
 - [63] Timothy A. Salthouse and Cecil L. Ellis. 1980. Determinants of Eye-Fixation Duration. *The American Journal of Psychology* 93, 2 (1980), 207–234. <http://www.jstor.org/stable/1422228>
 - [64] Monib Sediqi. 2023. Eye Tracking with Mediapipe. Retrieved Feb 2, 2024 from <https://kh-monib.medium.com/title-gaze-tracking-with-opencv-and-mediapipe-318ac0c9c2c3>
 - [65] Rashik Shrestha. 2023. Perspective-n-Point (PnP). Retrieved Dec 16, 2023 from <https://medium.com/@rashik.shrestha/perspective-n-point-pnp-f2c7dd4ef1ed>
 - [66] Ludwig Sidenmark and Hans Gellersen. 2019. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 27, 1, Article 4 (dec 2019), 40 pages. <https://doi.org/10.1145/3361218>
 - [67] Scott Stein. 2022. Watching Me, Watching You: How Eye Tracking Is Coming to VR and Beyond. Retrieved May 8 2022 from <https://www.cnet.com/tech/computing/watching-me-watching-you-how-eye-tracking-is-coming-to-vr-and-beyond/>
 - [68] Brian Jay Tang and Kang G. Shin. 2023. Eye-Shield: Real-Time Protection of Mobile Device Screen Information from Shoulder Surfing. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 5449–5466. <https://www.usenix.org/conference/usenixsecurity23/presentation/tang>
 - [69] TP-Link. 2023. AC1900 Wireless MU-MIMO Wi-Fi 5 Router. Retrieved Dec 16, 2023 from <https://www.tp-link.com/us/home-networking/wifi-router/archer-c80/>

- [70] Boston University. 2024. Wilcoxon Signed Rank Test. Retrieved Apr 29, 2024 from https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric6.html
- [71] Varjo. 2023. Varjo XR-4 Series. Retrieved Dec 16, 2023 from <https://varjo.com/products/xr-4/>
- [72] Jonathan Vaughan and Thomas M. Graefe. 1977. Delay of stimulus presentation after the saccade in visual search. *Perception & Psychophysics* 22, 2 (01 Mar 1977), 201–205. <https://doi.org/10.3758/BF03198755>
- [73] Wang, Sung, and Ronda Venkateswarlu. 2003. Eye gaze estimation from a single image of one eye. In *Proceedings Ninth IEEE International Conference on Computer Vision*. 136–143 vol.1. <https://doi.org/10.1109/ICCV.2003.1238328>
- [74] A. L. Yarbus. 1967. *Eye movements and Vision*. Springer.