XSec Companions: Exploring the Design of Cybersecurity Companions

Sarah Delgado Rodriguez
University of the Bundeswehr Munich
LMU Munich, Germany
0000-0002-0514-9846

Sarah Elisabeth Winterberg
LMU Munich, Germany

Franziska Bumiller
University of the Bundeswehr Munich
LMU Munich, Germany
0009-0003-1712-2039

Felix Dietz
University of the Bundeswehr Munich
LMU Munich, Germany
0000-0002-0241-0295

Florian Alt
LMU Munich, Germany
0000-0001-8354-2195

Mariam Hassib
RWTH Aachen, Germany
0000-0001-6530-9357

***Abstract*—Cyberattacks frequently target humans, for example, by using social engineering to trick them into revealing sensitive information or by exploiting insecure behavior. Traditional security awareness training and guidelines have proven insufficient to address this issue, as they are not tailored to individual usage conditions and are disconnected from real-world situations. Additionally, these trainings and guidelines do not adapt to users' changing needs or evolving knowledge. Contrary to traditional training, personal cybersecurity companions, whether digital or tangible, provide a new, adaptive, and integrated way to assist users in understanding security concerns and behaving securely in cyberspace. In this paper, we explore the space of cybersecurity companions through ideation workshops ($N = 12$), particularly focused on privacy in IoT and phishing. Through the analyses of the end-user visions built during our workshops, we conceptualize and present the XSec Companions design framework. Our work can guide future researchers in developing both digital and tangible xSec Companions whilst providing an overview of the opportunities and challenges in this space.**

***Index Terms*—design, explainable security, companions**

I. Introduction

Cybercrime has caused trillions of euros in global damage annually in recent years and is expected to grow by about 15% per year [1]. Up to 95% of incidents result directly or indirectly from user-centered attacks, where impostors exploit human factors such as limited knowledge or inattention [2]. Common examples include entering credentials on phishing sites, reusing passwords (enabling credential stuffing), or skipping VPN use on unencrypted networks.

As a result, *effective cybersecurity training* has emerged as a critical issue. While text-based guidelines and videos remain widespread, recent work suggests that interactive, context-aware approaches yield better long-term adherence [3]. Many existing programs lack sound theoretical grounding and fail to produce lasting changes in user behavior [4]. Researchers therefore advocate habit-creation strategies: simplifying training materials and identifying cues that trigger secure routines can improve user commitment [5]. Similarly, embedding cybersecurity education into daily life through personalization and engaging tasks shows strong potential [6].

Interactive digital and tangible companions are a novel way to deliver cybersecurity guidance more engagingly, intuitively, and in a user-centered manner [7]–[10]. In this work, we define *cybersecurity companions* as *(intelligent) agents that use human language to assist users in cybersecurity tasks and topics*. Inspired by virtual pets (e.g., Tamagotchi, Furby, Eilik[1]), these companions offer real-time assistance, feedback, and encouragement through language-based interaction [7]–[9]. By weaving playful engagement into everyday routines, they aim to shift security from a tedious chore to an interactive and more enjoyable part of daily life.

Although multiple projects have produced tangible and, more frequently, digital instantiations of such companions (particularly chatbot-based), a theoretical foundation for designing these systems remains underexplored. In particular, related works offer limited guidance and theoretical insights on how to design these companions effectively. To address this gap, we conducted ideation workshops ($N = 12$) and derived the *xSec Companion Design Framework*. Grounded in the Explainable Security (xSec) paradigm by Viganò and Magazzeni [11], our framework distills practical insights for the "why," "how," "who," "what," "where," and "when" of designing cybersecurity companions. The xSec paradigm emphasizes providing clear, context-aware explanations of security decisions, which is critical to ensure that users can trust and understand the provided explanations of cybersecurity. By rooting our design framework in the xSec paradigm, we aim to ensure that cybersecurity companions not only guide users but also foster trust and a deeper understanding of the rationale behind security practices, making these behaviors more meaningful and easier to adopt.

We finally illustrate how our design framework can inform new xSec Companions while discussing open challenges and opportunities in this evolving field.

Contribution Statement. Our work contributes to human-centered security by exploring the design of cybersecurity

[1]https://energizelab.com/consumerview/eilik, last accessed Jan. 18, 2025

companions. Specifically, we (a) conducted ideation workshops with 12 participants to elicit design requirements and visions, (b) synthesized these insights into the *XSec Companion* design framework, and (c) discussed its applicability by showcasing selected companions from related work and deriving future research opportunities and challenges.

II. Related Work

Our paper draws from related works on tangible and digital interfaces that provide cybersecurity assistance for users.

A. Companion-Like Cybersecurity Agents

Related work has explored tangible and digital companion-like agents to support secure user behavior. Chiou et al. [7] used interactive storytelling with a Zenbo robot to teach K-12 students about phishing and password safety. Follow-up work with teachers found the story-based model improved comprehension and attention [12]. Similarly, Pasquali et al. [8] used a Furhat tabletop robot to provide adult users with advice during gamified cybersecurity scenarios, showing how affective cues can influence decisions. Guo et al. [13] introduced a voice agent that emphatically discusses potential online payment fraud. Pears et al. [14] transformed case-based cybersecurity learning into an interactive chatbot for medical workers, showing positive engagement and usability. Adinolf et al. [15] ran expert focus groups to conceptualize a VR-based virtual agent, finding preferences for robot- or animal-like companions across thematic, stylistic, and mechanic categories.

These works show that interactive agents – embodied or virtual – can improve user engagement and cybersecurity learning. However, a broader perspective on how such companions should look and act remains lacking.

B. Novel Security Warning Interfaces

Researchers have also explored novel warning designs for immediate threats. Minakawa and Takada [16] used "kawaii" (a Japanese notion of "cute") graphics with animation and sound to counter warning fatigue. Silic [17] introduced "Maya," a chatbot with a humanoid avatar that engaged users when facing insecure websites, improving adherence through proactive dialogue. Bravo-Lillo et al. [18] found that *blocking actions* and added interaction reduced habituation to warnings.

Beyond visual and auditory cues, tangible and haptic feedback have also been used. De Luca et al. [19] employed lights and vibration in a "moody" keyboard to signal threats. Napoli et al. [20] added thermal feedback via a laptop heating pad. Do et al. [21] designed a wristband using squeezing sensations to alert users. While creatively engaging, these systems often remain one-off solutions rather than persistent companions.

C. Summary

Related works show that (1) *social or companion-like agents* can enhance user engagement and potentially improve adherence to secure computing practices, and (2) *novel warning mechanisms* can help overcome habituation by capturing users' attention at critical moments. Moreover, (3) related works usually focus on individual implementations – such as robotic storytelling or specialized chatbots – but do not conceptualize the design of cybersecurity companions more broadly [7], [8], [12], [13], [16], [17], [19]. To address this gap, we collect end-users' visions of cybersecurity companions through ideation workshops, interpreting them through the lens of the *Explainable Security (xSec)* paradigm [11].

Our goal is to (1) shed light on how users envision cybersecurity companions (e.g., for phishing and IoT privacy), (2) explore "why," "how," "who," "what," "where," and "when" these companions offer assistance, and (3) derive a framework from our insights to guide future designs.

III. Research Approach

A. Defining Cybersecurity Companions

Companions (aka artificial/social companions [10], [22], [23] or social/virtual agents [9]) shift user interaction from one-off or purely functional to relationship-based experiences. Benyon and Mival [9] define companions as *"intelligent, persistent, personalized, multimodal interfaces"* that *"change interactions into relationships"* [9, p.1], fostering *"an accessible, pleasing relationship with an interactive source in which there has been placed a social and emotional investment"* [9, p.1]. While many serve *entertainment* purposes (e.g., Tamagotchi), others fulfill *informative*, *educational*, or *caring* roles (e.g., tutors or medical assistants). Danilava et al. [23] emphasize language-based aspects, defining *artificial conversational companions* as *"computer agents that simulate human language behavior, and are aimed to serve, to assist and to accompany their owner over a long period of time"* [23, p.1]. They typically (a) possess conversational skills (cognitive, emotional, sociocultural, natural-language processing), (b) adapt to individual needs, (c) offer practical value, and (d) support long-term engagement. While these definitions highlight social presence, personalization, and relationship building – *cybersecurity companions* add domain-specific functionality: guiding users in adopting and maintaining secure behaviors.

Definition. A *cybersecurity companion* is an (intelligent) agent that uses human language to assist users with cybersecurity tasks and topics. It can be informative and/or educational, have an avatar, and be either digital (virtual) or tangible (physical).

B. Companions & Secure Behavior

Sasse et al. [24] present the *Security Behaviour Curve* to explain how organizations can foster lasting secure behavior among employees. Traditional cybersecurity training often focuses on providing *information*, raising *awareness*, and ensuring basic *understanding*. However, Sasse et al. argue that five additional steps are needed for long-term impact.

The first is *concordance*, where employees must willingly commit to behavior change. We think that cybersecurity companions could support the creation of concordance by translating abstract concepts into accessible, context-relevant explanations. By enhancing users' understanding of cybersecurity risks and secure behaviors, companions can help

them recognize the (personal) benefits of behaving securely, making it easier for them to align with secure behavior goals. Next, Sasse et al. [24] discuss employees' belief in their own *self-efficacy* – confidence in executing secure practices. In our opinion, companions could offer in-situ guidance, feedback, and encouragement, lowering entry barriers and boosting users' confidence. The third step Sasse et al. [24] discuss *implementation*, requires that employees have the skills and enabling environments (e.g., usable tools and simplified processes). Companions could support this by directing users to available tools and guiding them through the processes. Repeated secure behavior then becomes *embedded* into routines – supported by nudges and forgetting insecure habits. By being persistently present in users' daily routines, companions could deliver timely nudges and reinforcement, supporting the internalization of secure behavior.

C. Research Questions

Building upon these considerations, our work explores the design of cybersecurity companions. It answers the questions:

RQ1 – How can cybersecurity companions be designed in terms of tangible or digital form, visual style, and integration with existing devices or platforms?

RQ2 – In which ways can cybersecurity companions assist users, considering both immediate risk detection (e.g., phishing alerts) and ongoing support (e.g., password hygiene, IoT device management)?

D. Methodology

To build a theoretical foundation around the design of these companions, we adopted a multi-stage empirical approach:

1) **Ideation Workshops:** We elicited users' visions of cybersecurity companions via participatory workshops. Participants enaged with how such systems should appear, function, and integrate with daily technology.
2) **Framework Synthesis:** Drawing on the Explainable Security (xSec) paradigm, we used thematic analysis to synthesize core design dimensions ("why," "how," "who," "what," "where," and "when") into a cohesive *XSec Companion* framework.
3) **Discussion on Applicability:** We then applied our framework to selected prior research, illustrating how it applies to and unifies scattered insights and might guide the design of new companion systems.

Our approach balances participatory, user-centric exploration with a conceptual lens (i.e., the xSec paradigm and related work) to organize and interpret the findings. By grounding our design framework in xSec, we aim to ensure that the resulting support provided by xSec companions is comprehensible and trustworthy for users.

IV. Ideation Workshops

Related work has proposed directly involving end users in the design of novel cybersecurity interfaces through participatory design, rather than relying solely on feedback to developer-created proposals [25]–[28]. This approach can yield deeper insights, be more efficient than iterative feedback loops, and lead to more user-friendly systems [25], [28]. Integrating end users into the design process may even be essential for a truly human-centered perspective when reimagining cybersecurity to achieve security by design [26].

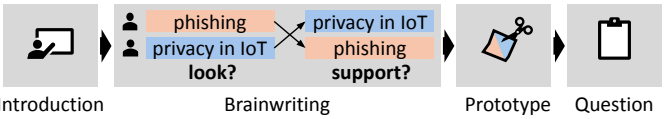


Fig. 1: The workshops ($N = 12$) consisted of (a) an introduction, (b) a brainwriting task about how XSec companions could look like and assist their users, (c) a paper-prototype crafting session, and (d) a final questionnaire.

As our goal was to elicit potential design dimensions of cybersecurity companions, we conducted participatory ideation workshops. Our workshop format was inspired by Andersen and Wakkary's [29] "The Magic Machine Workshops" to encourage open-ended exploration of cybersecurity companions without biasing participants toward digital or physical implementations. To support broader idea generation, we added a preliminary brainwriting task before introducing material constraints. This task was guided by the 635 brainwriting method [30] that encourages the generation of ideas in a structured, non-hierarchical, and collaborative way, allowing all participants to contribute without the inhibition often caused by more vocal individuals. This enabled participants to collaboratively focus on the functional and experiential aspects of cybersecurity support rather than predefined technological forms.

To answer our research questions, we organized three ideation workshops ($N = 12$) to collect end-user design ideas for cybersecurity companions addressing *phishing* and *IoT privacy*. We selected these two topics because both are central to today's cybersecurity landscape. Phishing remains one of the most widespread cyberattacks, affecting individuals and organizations worldwide [31]. At the same time, the proliferation of IoT devices has introduced privacy challenges that many users are only beginning to understand. Focusing on these distinct yet high-impact areas allowed us to capture both well-known attack vectors (phishing) and emerging, technology-driven concerns (IoT privacy), thereby reflecting a broad spectrum of real-world user needs.

A. Recruitment & Registration

We recruited participants through our institution's social media channels and mailing lists, providing a link to an online questionnaire which included (1) a consent form detailing study procedures, (2) basic demographic questions, and (3) several possible workshop dates. We also asked each participant to provide a nickname for use during the workshop sessions, allowing them to withhold their real names.

B. Study Procedure (see Figure 1)

The workshops lasted 90 minutes, with each participant being involved for the full duration.

1) Introduction: We first greeted participants and asked them to introduce themselves using their nicknames to break the ice. After confirming consent, we started audio and video recording and briefly outlined the study's goals. A short presentation covered phishing, smart home privacy (see Appendix A), and that companions could be *digital* or *tangible*. Drawing on Jung et al. [32], we discussed that avatars can range from god-like to plants, animals, or objects. We then showed examples (see Appendix C), including the animated dog from [16], Bixby, the CiSA chatbot, Cortana, humanoid avatars [33], the "Privacy Flower" [34], Siri, a Tamagotchi, and the Zenbo Robot [7], to broaden views on what a "companion" can be.

2) Brainwriting: To spark ideas, we showed a slide with companion images and divided participants into pairs or groups of three. Using a simplified 635 brainwriting method [30], [35], each participant wrote down ideas on the *appearance* of a companion for *phishing* or *IoT privacy*, then exchanged notes to *expand* or *adapt* them. They next considered *how companions might assist* users. Groups placed their notes on a flipchart and briefly presented their concepts.

3) Prototype Crafting: Participants concretized their ideas by developing prototypes with basic crafting materials either for "phishing" or "IoT privacy" (30 minutes). We asked them to consider (a) the companion's *look and feel* (i.e., appearance and materials [36]), (b) its *core functions*, and (c) how it *supports* users . Groups presented their prototypes' names, purposes, and interaction styles.

4) Final Questionnaire: We ended with a short questionnaire asking participants to choose their favorite prototypes and explain why. This provided insight into their values and expectations for an *ideal* cybersecurity companion.

C. Collected Data & Analysis

We collected audio/video recordings of the workshops and all participant-generated materials (i.e., sticky notes, prototypes, and questionnaires). One researcher transcribed the recordings, digitized all handwritten notes, and grouped participant quotes with corresponding notes and prototypes. Questionnaire responses were also digitized and organized by question. A second researcher reviewed the dataset for accuracy. Both researchers then discussed all results to build a shared understanding. Following Naaem et al. [37] and Fife and Gossner [38], we conducted a hybrid thematic analysis (deductive and inductive), grounded in the xSec paradigm [11]. As per Braun and Clark [39], [40], we did not code independently but iteratively refined our mutual observations.

D. Ethical Considerations

Our institution's IRB reviewed and approved the study protocol. We provided detailed information about the study's goals and procedures to potential participants before registration. All participants gave written consent for data collection and were informed when audio and video recordings would begin and end. We also requested that they use nicknames instead of real names to maintain anonymity. Participants received information on their rights as subjects, including how to contact the research team with questions or concerns.

TABLE I: We conducted three workshops with three to 5 participants in each workshop (WS).

	overall	**WS 1**	**WS 2**	**WS 3**
N (groups)	12	4 (G1, G2)	5 (G3, G4)	3 (G5)
gender	9M & 3F	2M & 2F	4M & 1F	3M
age range	23 - 35	25 - 35	23 - 24	26 - 29
age mean (std)	26.2 (3.7)	29.0 (4.6)	23.2 (0.5)	27.3 (1.5)

TABLE II: Ideas for cybersecurity companions suggested by participants during the presentation of brainwriting results.

Group	**Phishing**	**Privacy in IoT**
G1	cute animal, anthropomorphized envelope	historical figure, calender, physical cover
G2	human authority figure, anthropomorphized symbol	guard, protective pet
G3	genie, pop-up warning sign	robot, talking portray, ball
G4	fishing rod or fish symbol	physical shutter, hardware mute button, red light
G5	favorite movie or anime character, god-like chatbot	physical voice companion, app-based voice companion

E. Limitations

Our findings may be subject to social desirability bias and availability bias since we showed participants examples of existing social companions. Nevertheless, we made sure to show a broad variety of companions that were also composed of varying materials. Moreover, our participant sample consisted of young and predominantly male students. As young university students, our participants can provide particularly valuable input because they could be among the first to actually experience cybersecurity companions in their daily lives. Nevertheless, the impact of the overrepresentation of male participants remains unclear. Hence, we invite future researchers to add to our findings by focusing on other populations. Finally, our workshops focused on phishing and privacy in IoT. While we showcase in Section VI-G that our findings can be applied to other cybersecurity topics, future work is needed to provide empirical proof.

V. Results

A. Participants

Table I shows the demographics of our participants and their distribution over the workshops. Overall, three of the 12 participants identified as female and nine as male. The ages of the participants ranged from 23 years to 35 years ($mean = 26.17$, $std = 3.66$). All participants were university students – seven of them also worked.

B. Brainwriting

1) Phishing:

a) Group 1 (G1): G1 proposed a *cute animal* companion integrated into the (web-)mail client. It is a *"companion which you may want to interact with because it's cute [...] and it's not so technical"* (P1). It also features a red exclamation mark and makes *"sounds like the animal it looks like, whenever there is a phishing attack it suspects"* (P2). They also suggested an *anthropomorphized envelope* to symbolize e-mails: *"So think of Clippy [...], but instead of a paper clip, it's a mail envelope to symbolize: hey it's about emails and Phishing"* (P1). If a phishing mail is detected, it would *"resist opening up [...][so users have to] click 3 times or 6 times"* (P2).

b) Group 2 (G2): G2 proposed a *human authority figure – "in a suit or something"* (P3) – that gives *"the user some lessons [...] to tell them: Okay, when an e-mail arrives, then please be careful"* (P4). They also suggested an anthropomorphized *symbol – "like a warning sign or an exclamation mark"* (P3) – which *"scans the e-mails and [...] marks [suspicious ones] with an exclamation mark and makes a red banner [...] so people won't just click on any links"* (P4). G2 added that the companion should be *cute*, use *bright colors to capture attention*, or appear as a *"pop-up [...] that blocks the text of the phishing e-mail"* (P3).

c) Group 3 (G3): G3 suggested a companion that looks *"like a genie or something to let the user feel a bit safer because a genie is somebody who knows everything"* (P6) or a *pop-up warning sign*. G3's companion would *"create like a [...] phishing register with scam links"*(P6).

d) Group 4 (G4): G4 suggested a *fishing rod or fish symbol* that appears if an email is suspicious to *"warn the user[and] [...] prevent him from getting phished"* (P9).

e) Group 5 (G5): G5 proposed avatars resembling the user's favorite *movie* or *anime character* – a *"human, so that you can also take it seriously"* (P12). They imagined this personalized companion would prompt users to *"have a quick glance and maybe smile [...] whenever it pops up and it says something"* (P12). They also suggested a *chatbot* that *"scans through all the documents you are reading [...], [identifies] potential risk[s] and [...] advises on what you should do"* (P11). This chatbot is described as *"god-like [...][and runs] in the background of a phone. With the help of AI, it can scan and test with some random usernames and passwords [...] [because] if [a login form] is fraudulent, then it will accept the [test] data"* (P11). Finally, G5 noted that phishing companions could *"detect some [...] suspicious [...] passages [...] and highlight them as a warning. [...] Such a companion could [...] raise the user's awareness [...] so the user can also detect [...] them themselves"* (P11) in the future.

2) Privacy in IoT:

a) Group 1: G1 suggested a *historical figure "which lived in times where privacy was quite important and not given [...] and which tells you what can happen when you really don't care about privacy"* (P1). G1 also mentioned a *calender* with *"reminders at regular intervals that the IoT device settings and privacy should be checked"* (P1) and a *geometrical* physical device that could *"block [a smart speaker] physically if we do not want to interact with it"* (P1).

b) Group 2: G2 proposed a *guard* or *protective pet* that *"can walk from device to device and check it"* (P4). As a dog, *"it could bark or alert when there are [...] intruders, people tempering with it or if there are illegitimate log-in attempts"* (P3). As a guard, it could *"only appear when the devices are recording without your consent [...] [or] when there are illegitimate log-in attempts"* (P3).

c) Group 3: "You can have like a robot, an image or portrayal of someone who is talking or an animal" (P6) which would *"ensure that [...] [IoT] system[s are] not taking any input, while [the users] don't want them to [...] [and in case] someone is trying to ping your IP that's not from your home network then you get [..] a warning"* (P7).

d) Group 4: G4 discussed a *physical camera shutter "that can be seen by the user for [indoor cameras]. So if it's shut, the user can not be recorded [...]. For outdoor cameras, it would be great to have a red light or something indicating, for example, to the postman that he's currently being recorded"* (P9). They also mentioned a *"hardware button for the microphone, so only if pressed can the user be recorded, for example, by Alexa"* (P9).

e) Group 5: G5 envisioned a *physical voice-based* companion – a *tangible object "that [users] can carry around and [...] talk to it, so that it tells [them] about potential risks regarding the security"* (P10) of nearby IoT objects. It *"automatic[ally] notifie[s its user] once something risky or shady is identified"* (P11). G5 emphasized the need for a centralized solution, noting they would not *"want ten thousand of companions [...] for each thing like [...] a companion for my toaster [...] and [...] my washing machine"* (P10). They also proposed a smartphone *app*-based, voice-controlled companion that is *"mobile and [...] applicable to many different things"* (P10) with an *"abstract [...] [avatar] because we have many, many things like Siri [...] so I feel that that's more natural than [...] talking to a pet or something"* (P10).

C. Crafted Prototypes (Photos in Appendix D)

1) Phishing:

a) Torpedo (G2): G2 called their prototype *Torpedo "because you can't fish when [...][there is a] Torpedo"* (P3). Torpedo is a virtual pop-up that is *"specifically designed for a workplace setting"* (P3). It *"pop[s] up and block[s] the text of a phishing e-mail"* (P3). Its design is intended to gain the user's attention through a *"cute little mascot creature, that has a very alert and surprised facial expression"* (P3). The prototype was created by combining white and red paper. It shows the avatar, red warning symbols, *"Warning! A potential phishing attack was detected."*) and instructions like *"do not click on any link [...] and contact your IT support"* (P3).

b) Snoopy I (G3): "Snoopy the First" (P6) is a *"big cat with a big warning"* (P7). Snoopy I flags *"sus[picious] mail[s], like 'buy my car for a dollar'"* (P6) and warns the user. It also *"add[s] the phishing mail into [the] phishing register and deletes it"* (P6). To show this, G3 sketched the Gmail inbox. *Snoopy I* is represented as a cat, with a speech bubble stating *'Attention!! Snoopy smells a phish..."*.

c) Safory (G5): Safory (SAFe accessORY) is a *"fashion accessory"* (P10) which users *"would like to carry around"* (P10). It can be worn as *"[a] watch, [a] necklace and so on"* (P10). The crafted prototype consists of a blue bottle cap with a string and an attached paper clip. G5 also drew a red cross sign and wrote the word *"Phish"* underneath. When the user receives a suspicious e-mail, Safory will *"project out [a] symbol [with a] cross"* (P12) onto *"whatever is in front of [the user], maybe the next person or the wall"* (P12) so it is like *"raising a red flag in front of [the user]"* (P12). In contrast to a screen on a necklace, which users might not always see the projection is *"wherever you look"* (P12). Safory *"will also vibrate [...] to catch your attention"* (P10). It incorporates a microphone *"so you can ask [it] questions"* (P12), and a camera to detect the user's surroundings.

d) Catch It (G5): Catch It is a browser extension similar to *"adblock [...] but it's for phishing"* (P11). G5 drew a sketch on paper, which shows how it scans an e-mail for suspicious aspects. Catch It scans incoming e-mails for specific keywords, compares the sender's mail address to a known actual e-mail address, and does the same for URLs in the e-mail.

2) Privacy in IoT:

a) Orwelly (G1): Orwelly is a physical warning device that informs its users about nearby IoT devices. P1 described it as *"a heart-shaped IoT Privacy device with many eyes and [...] a jammed-in little [...] object in one of them"*. Hence, G1 used a heart-shaped plastic piece with many googly eyes. One of the eyes has a wooden stick stuck in it. P2 added that Orwelly could shake and wiggle its many eyes to indicate that there is, for example, a camera recording the user. Hence, *"being recorded becomes a conscious choice [...] and if you're not okay with it, you can confront the owner [of the IoT device] or something"* (P1). Orwelly is *"very weird"* and *"very silly"* and thus *"grabs the attention immediately"* (P1): *"[I]t's not just a black box, which [...] you may ignore"* (P1).

b) Proofy01 (G4): G4's physical companion is *"called Proofy [...] because it comes with this [...] sound-proof and camera-proof"* (P8) box. The second object is smart speaker with a screen that can show a *"face or not, depending on [the user's] preferences"* (P8). "So [the smart speaker] has like its' little house (i.e., the box)], that he goes into and the user is kind of assured like okay, now if he's in that thing, you won't be recorded or anything" (P8). In addition, Proofy01 has a red LED light to indicate if it is currently recording. The crafted prototype consists of a cardboard box and a bottle cap with googly-eyed and red fluff on top.

c) Secure+ (G5): Secure+ is an app that integrates all the *"smart appliance devices, like your light bulb, your fridge or your key [...] to your house[.] [...] So all your house appliances are integrated into one device and constantly more monitored."* (P11). To show this G5 drew different widgets on a smartphone that represent a *"Key"*, a *"smart fridge"*, *"WIZ"* smart lights, *"IoT Security"*, or *"apple home"*.

D. Questionnaire

1) Preferred Companion: In the first workshop, 3 of 4 participants preferred *Orwelly*. P2 explained *"[i]t is silly[, so] it grabs attention. It is weird[,] so [it] makes you question why it is there and [...] what the purpose of the cameras recording you is"* (P2). P3 and P4 appreciated its symbolism: *"the idea of using a lot of eyes to tell the user that it is being watched is great"* (P4). P1 favored *Torpedo* for its *"cute mascot"*: *"While users may ignore typical warnings [...], this message with an unexpected mascot may give them pause (for a while)"* (P1).

In the second workshop, 4 of 5 participants liked *Proofy01* for its simple, effective privacy protection. P7 wrote *"any sort of physical cover ensures that no video or audio is being recorded"* and its *"manual setup is easy and you don't need to be a techie"* (P7). One participant chose *Snoopy I*, noting they do not use IoT devices.

All three participants in the third workshop chose *Catch It*. P11 and P12 valued its familiarity with Adblock: *"it is quite straightforward, building on top of what's existed already"* (P11). P10 noted that a *"browser extension will reach more potential users [...] compared to [...] buying something"*.

2) Positive Aspects: 6 participants saw a *positive effect of cybersecurity companions on their risk awareness*. 5 participants also described that such companions would make them feel more secure. Hence, P12 wrote *"more awareness, less paranoia about [protecting] your privacy"*. 4 participants also felt that such companions could be *more effective than other methods*. P4 argued that *"as they are constantly present, they might be more efficient, than having a tutorial every year instead (at work for example)"*. 3 participants mentioned that cybersecurity companions are *fun to use*. P1 mentioned that they *"gameify IT Security"* and *"make it more fun"*. 3 participants also found that cybersecurity companions are *easy to understand and use*. Hence, P7 argued that *"it does not give out warnings that are unintelligible to someone who might not know a lot about tech"*. Moreover, P8 wrote that they *"can be tailored to suit different types of understanding in levels of technology[,] [...] age groups, cultures, etc."*.

3) Negative Aspects: 4 participants mentioned that cybersecurity companions could *become a threat themselves*. P5 mentioned that it *"can be a vector for online scams"* and P12 wrote that if it *"gets attacked, then you will trust [it] blindly"*. 3 participants were concerned that *cybersecurity companions could malfunction*. 3 participants also mentioned that such companions could *make people uncomfortable*. P2 wrote that if they *"startle or shock, [they] might make someone uncomfortable"*. P4 added that *"people might feel uncomfortable 'to be watched' how they work on their computers at work"*. 3 participants were also worried that *users could ignore the companion.* Hence, P4 argued that *"people might ignore them because they take them not as seriously"* and P3 wrote that *"if their design is too cute or familiar, they might normalize the privacy threats they are supposed to protect from"*. Two participants expressed *privacy concerns*, arguing that they pose *"a potential risk for your privacy"* [P6]. Furthermore, 3 participants indicated they could create a *false sense of*

security or overdependence on them. Finally, P1 mentioned that physical companions could *create more waste*.

4) Potential Adoption: We finally asked participants whether they would see themselves using a fully functional version of any of the prototypes. Participants' answers ranged from 2 (disagree) to 5 (strongly agree). The median of participants' answers is 4 (i.e., agree), indicating that most participants would potentially adopt a cybersecurity assistant.

E. Summary: Addressing the Research Questions

While participants generated a broad spectrum of ideas, not all fit the definition of *cybersecurity companions* outlined in Section III-A. In particular, concepts that offered purely physical barriers (e.g., camera shutters, mute buttons, or simple visual indicators) without any language-based interactions were excluded from our analysis and framework derivation.

a) RQ1 – Appearance, Materiality, and Integration: Participants envisioned both *tangible* and *digital* companions. Tangible companions were especially popular for IoT privacy scenarios, often conceived as external add-ons or standalone systems (e.g., a wearable or moving robot). In contrast, digital companions – primarily for phishing – were typically integrated into email clients or browsers. The companions featured human-, animal-, or object-like avatars to foster user engagement. This variety underscores participants' desire for companions that are either unobtrusive integrations into existing ecosystems or playful physical entities.

b) RQ2 – Modes of Assistance: Participants consistently highlighted companions' roles in *warning* users about imminent threats (e.g., suspicious emails), *educating* them on best practices (e.g., password hygiene), and responding to questions *on-demand* (e.g., clarifying whether specific links or IoT configurations are risky). Many designs involved data analysis features – scanning inbox contents or checking device configurations – to deliver targeted assistance. These findings demonstrate the potential for cybersecurity companions to provide both proactive alerts and on-demand expertise, thereby bridging the gap between general security guidelines and context-aware interventions.

VI. The xSec Companion Design Framework

Grounded in the Explainable Security (XSec) paradigm [11], we performed a hybrid thematic analysis to derive the xSec Companion Design Framework (see Figure 2).

A. ***Why*** *do we need companions?*

Participants consistently noted that XSec Companions would make them *feel more secure* and increase their *awareness* of cybersecurity risks. They considered language-based interaction highly *intuitive*, allowing even non-experts to communicate with the system easily. Moreover, participants described these companions as *more engaging* than traditional awareness training platforms. This leads to stronger *emotional reactions* and a sense of enjoyment – particularly when the companion featured playful or visually appealing avatars.

B. ***How*** *do the companions assist?*

Participants envisioned different levels of assistance that XSec Companions could provide. At a *general* level, companions would focus on *education* – explaining the importance of online privacy or illustrating secure behaviors, as with G1's historical figure or G2's authoritative agent. In more *context-specific* scenarios, they might *highlight* risky elements (e.g., suspicious links, emails, or devices) to prompt behavioral adjustments or deliver *warnings* about threats and best practices (e.g., Torpedo, Snoopy I). Some concepts also involved *blocking* harmful actions – such as opening phishing emails – or *blocking* data collection at the system level, as seen in the Proofy01 prototype. This layered approach suggests XSec Companions can combine educational and protective roles, tailored to users' needs and contexts.

Note that Gundu [41] offers a relevant overview of ChatGPT-driven chatbots assisting users with cybersecurity.

C. ***Who*** *provides the assistance?*

Participants envisioned a broad range of XSec Companions, including *(anthropomorphized) symbols or objects, animals, humans, and even god-like forms* [32]. Across these diverse concepts, several common themes emerged. First, companions should *exude authority* to foster user trust – often through professional or confident visual cues. Second, they should possess an *engaging appearance*: participants emphasized the value of personalized or cute designs that *catch the user's eye* with bright colors, movement (e.g., animated facial expressions), and playful surprises. Third, companions could be fully *digital* (e.g., software agents) or *tangible* (e.g., physical devices), reflecting the desire for flexible integration into different environments.

D. ***What*** *is explained?*

Our participants proposed XSec Companions capable of delivering both *general cybersecurity information* and *context-specific insights*. For instance, a historical-figure companion (G1) could explain why privacy matters, while others focused on *particular risks* related to technologies like email or IoT devices. Beyond simply highlighting hazards, many prototypes also surfaced *reasons why a user's security could be in jeopardy* – for example, by marking suspicious passages in potential phishing emails. This explanatory detail helps users understand the underlying threat patterns, potentially enabling them to recognize similar red flags without the companion's assistance in the future. Several participants proposed that companions offer *step-by-step instructions* for more secure behavior, further emphasizing the educational and empowering role that XSec Companions can play.

E. ***Where*** *is the assistance provided?*

Participants envisioned different approaches to *where* XSec Companions could deliver assistance. Some proposed *independent* companions (e.g., G1's historical figure) operating outside specific platforms, others focused on *single-system* support (e.g., Proofy01's smart speaker), and some imagined

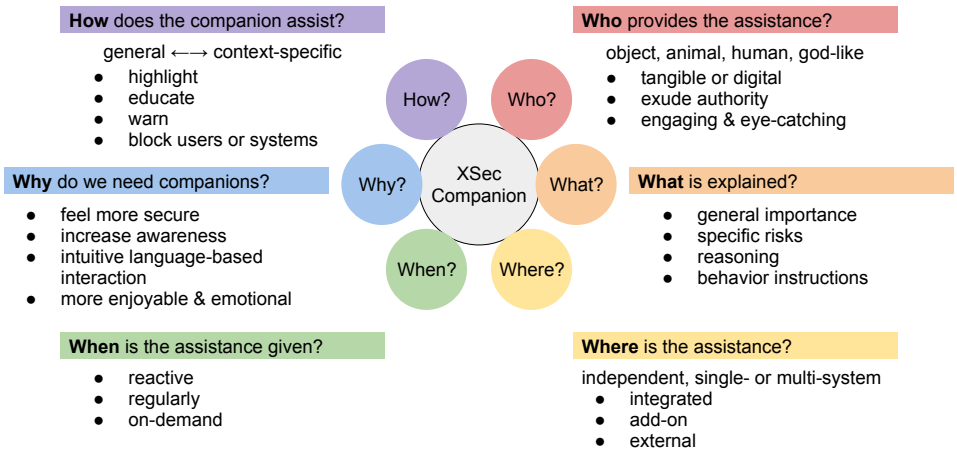


Fig. 2: The xSec Companion Design Framework synthesizes our findings on potential appearances and functionalities of cybersecurity companions. In particular, it answers the questions: "who?", "what?", "where?", "when?", "how?" and "why?".

multi-system designs (e.g., G2's protective pet) for centralized oversight of various devices or software. In terms of integration, both digital and tangible companions could be *embedded* in existing systems – like email clients (e.g., Snoopy I) or hardware (e.g., Proofy01 with anthropomorphic features and LED indicators). These companions enhance familiar interfaces while offering interactive features such as warnings, prompts, or notifications. Alternatively, companions might act as *add-ons* (e.g., Catch It as a browser extension or the Proofy01 box), enriching existing tools without full integration. Participants also proposed *external* companions offering cross-system support. For instance, G5's chatbot scans all open apps for threats, while Safory projects warnings into the user's environment. These solutions enable centralized, flexible oversight beyond the limitations of single devices.

F. *When* is the assistance given?

We also derived multiple possibilities of when an XSec companion can assist. In particular, most companions proposed by our participants *reacted to potential threats*. For example, Torpedo is a pop-up that occludes potential phishing emails, and Safory projects warnings onto the user's surroundings. Another option is to provide the user with information regularly. For example, G2's human authority figure companion educates its users about secure behavior from time to time, and G1's calendar companion regularly reminds its users to check their IoT devices' privacy configurations. Finally, xSec Companions can also provide *on-demand* assistance without reacting to risks or temporary triggers. Users can freely choose when to interact with those types of xSec Companions, making their functionalities completely dependent on the users' individual needs. For example, G5 suggested a tangible object that users carry around and ask for cybersecurity advice whenever they want, and the Safory wearable has similar capabilities.

G. Application to Companions from Related Work

We discuss a variety of cybersecurity companions previously introduced in the literature to reflect on the applicability of our proposed design framework. To ensure a meaningful sample, we performed a three-round snowball-sampling literature search [42]. We started with eight publications [7], [8], [13]–[17], [43] that initially inspired our work. One researcher iteratively screened the titles – and, if needed, abstracts – of both the references within these eight papers and the works citing them that were published until January 16, 2025. In total, 927 titles or abstracts were reviewed based on our companion definition from Section III-A, resulting in 40 relevant publications (Appendix E). However, 33 of the 40 publications were chatbot-based xSec Companions. Since our aim here is to discuss the applicability of our Design Framework on diverse types of companions, we present only one of the chatbot companions [17] in detail.

Chiou et al. [7] developed *regular* storytelling-based classes for children about safety, security, and privacy using a Zenbo *robot teacher*. Students could choose between different action options at certain points of the story. Hence, their approach made use of *intuitive language-based interactions* to *educate* users on *general* cybersecurity topics. The authors chose an *external tangible social robot* to deliver the educational stories because interacting with them is *enjoyable* for the children, which increases *engagement*. The stories do not focus on risks specific to certain technologies, but *generally transmit the importance of privacy, security, and safety*.

Pasquali et al. [8] implemented a social-engineering-themed *educational* game where adult users received *spoken assistance* from a connected *tabletop robot head*, the Furhat robot. In particular, the *robot suggested how users should behave in specific situations in the game* and provided sometimes *reasons for their recommendation*. However, users were more affected by *emotional cues* (i.e., facial expressions and expressions of worry) than by logical reasoning.

TABLE III: We show how cybersecurity companions from related work apply to the xSec Companion Design Framework.

Companion	Why?	How?	Who?	What?	Where?	When?
Robot Teacher [7]	intuitive language-based interaction & enjoyable	general education	engaging tangible Zenbo robot	general importance	external independent assistance	regularly
Tabletop Robot [8]	intuitive language-based interaction & emotional	context-specific warnings	tangible Furhat robot	(reasoning &) behavior instructions	single-system add-on	reactive
VR Agent [15]	increase awareness, enjoyable & emotional	context-specific education	engaging digital cyborg, robot or animal in VR	specific risks	external system-independent assistance	– not disclosed –
Maya [17]	increase awareness & intuitive language-based interaction	context-specific education	digital female human	behavior instructions	single-system add-on	reactive
Shing [13]	intuitive language-based interaction & emotional	context-specific warning	engaging digital agent, no visual avatar	behavior instructions	external single-system assistance	reactive
"Kawaii" Warning [16]	increase awareness & emotions	context-specific warning	eye-catching digital dog	specific risk	integrated single-system	reactive

Adinolf et al. [15] proposed using gamified *VR experiences with digital agents* for *cybersecurity training*. They argue that using agents for education imitates the *emotional connection* that users might have with human teachers or classmates, making education *more enjoyable* as opposed to digital education tools without agents. The authors conducted ideation workshops and found that participants envisioned *cyborgs, robots, or animal avatars* that guide users through *specific tasks* like identifying malicious emails.

Silic [17] developed *Maya, a chatbot with a female human avatar* that was *added to* the insecure website *warning* in Google Chrome v42 to reduce habituation and *increase awareness*. Maya encouraged users to *chat about their intentions* and offered advice on how to verify if a site was malicious, ultimately *educating* them on *secure behavior* in this *specific context*. Similar xSec Companion chatbots have been developed for phishing [43]–[45], privacy settings in social media [46], [47], and security training [48], [49].

Guo et al. [13] developed *Shing* an empathic conversational voice agent that performs phone calls to potential online payment fraud victims. Shing *reactively warns* its users about suspicious transactions and *instructs them to cancel transactions* if the suspicions are confirmed during the phone call. As Shing was more effective than a similar non-empathic voice agent, the authors demonstrated an *emotional effect on users*.

Minakawa and Takada [16] augmented a *digital* insecure website *warning* with an animated *"kawaii"* dog avatar to reduce habituation, *increase awareness*, and enhance warning effectiveness by triggering *affective responses*. They found the augmented dialog more eye-catching than traditional warnings.

VII. Discussion

A. Developing xSec Companions with the Framework

We envision our framework guiding the design of future xSec companions. Although partly speculative, we key design considerations and reflections from related work here.

Developers may begin with the *"Why?"* dimension and our observations in Sections VII-B and VII-C, to assess whether xSec Companions suit their specific use case.

The *how, what,* and *when* of assistance depend on whether the goal is to *generally educate users* or address *specific threats*. A *context-specific* companion requires awareness of when and where risks occur (e.g., phishing, unsafe networks), enabling *reactive* assistance. While this would certainly make it much more useful, it would also introduce privacy and implementation complexities. An alternative is *on-demand assistance*, where context data is only collected after users request help. This preserves privacy but assumes users can identify risks and articulate questions (e.g., in chatbot conversations), making it suitable for more aware users. A key challenge is balancing autonomy with guidance. *Highlight-only* companions inform without intervening, supporting autonomy but risking neglect by novices. *Educating warnings* guide behavior and foster self-efficacy [24], [50], but can be disruptive and lead to *warning fatigue* [51] – especially if ill-timed (e.g., during focused tasks). More forceful companions may *block unsafe actions*, which limits autonomy and should be reserved for high-risk situations. In such cases, clear explanations and *override options* help preserve transparency and user control [24], [52].

The *who* (avatar) and *where* (physical or digital setting) of the companion are often intertwined. *Tangible* companions – especially in centralized or stationary setups – can feel more *engaging, emotional and harder to ignore* [7], [8], [12], though larger tangibles are be better suited for *stationary or centralized* setups and centralized assistance across multiple systems. Smaller ones could serve as *wearables* or *add-ons* (e.g., keychains or phone cases) [36]. In contrast, *digital* companions are easier to *integrate or add* to existing systems and often focus on a *single use case*, typically implemented as *chatbots* with or without avatars. Both types require thoughtful avatar design. While robotic and human forms were common in our workshops, participants also proposed

symbolic representations (e.g., pop-up blockers, fishing rods). Designers should consider how an avatar's appearance shapes expectations, trust, and mental models. *Anthropomorphized symbols* – combining human traits (e.g., eyes) with security elements (e.g., lock icons [53]) – may offer a compelling mix of approachability and relevance.

B. Opportunities for Future Research

1) Context-Specific, Multi-System xSec Companions: Most xSec Companions in prior work address *single* security problems, such as online-payment fraud or secure messaging, rather than spanning multiple security domains. For example, only the robot teacher by Chiou et al. [7] offered a more general cybersecurity focus. We also found no research on xSec Companions designed to handle *multiple* threats across different platforms while remaining context-aware and adaptable to users' immediate needs. Future work could explore *multi-system* companions that manage diverse risks – such as phishing, IoT privacy, and password security – while adjusting to changes in user contexts in real-time.

2) Integration of Artificial Intelligence: As large language models (LLMs) and AI tools become increasingly accessible, they present new possibilities for both threat detection and user-facing interactions [54]. LLM-enhanced systems could *personalize* advice based on behavioral patterns, *automate* advanced threat analysis, and conduct more *nuanced* conversations with users [41], [55]. While many recent publications leverage AI-driven, chatbot-like xSec Companions (see Appendix C), they typically exclude *embodied or multimodal designs*. Exploring how AI can seamlessly integrate with avatars or tangible elements is a promising direction to enrich user experiences and potentially improve effectiveness.

3) Building Custom Tangible xSec Companions: Existing tangible xSec Companions rely primarily on off-the-shelf robots [7], [8], [12]. While these studies highlight the value of physical embodiment for user engagement, there remains substantial potential for *custom hardware* tailored specifically for cybersecurity contexts. Such implementations would allow greater flexibility in appearance, interaction modalities, and functionality – ranging from compact wearables (e.g., like the Safory prototype) to fully integrated smart-home installations. By experimenting with different form factors, researchers could more thoroughly address issues of portability, usability, and ongoing maintenance, ultimately broadening the scope and practicality of tangible xSec Companions.

C. Challenges of xSec Companions

1) Avoiding User Annoyance: A well-known example of a digital companion failing to assist effectively is Microsoft Word's "Clippy," which lacked empathy, self-awareness, and adaptability [56], [57]. Designing xSec Companions to *genuinely assist* without distracting from users' tasks is challenging, especially since cybersecurity tasks (e.g., password management, phishing checks, privacy controls) can already feel overwhelming [58]–[62]. A promising direction is to create *emotionally intelligent* xSec Companions that *context-sensitively* adapt to users' needs, skill levels, and focus. For example, companions might delay alerts during high-focus tasks and avoid repetitive warnings. Workshop participants also noted that *customizable appearances* could reduce annoyance by fostering personal connection or enjoyment.

2) Planing for Errors & Failures: Designing xSec Companions also involves addressing potential risks that may arise from *incorrect advice or system failure* [63], [64]. As mentioned by our participants, the companion may put users in danger or cause unnecessary irritation if it misidentifies dangers. Companions should *openly express ambiguity* (e.g. "this might be suspicious") [64]and welcome user comments to correct or flag errors in order to lower this risk. Building confidence can also be facilitated by implementing a "second opinion" procedure, such as escalation to IT support.

3) Preserving User Autonomy: Depending on their design, xSec Companions can inadvertently erode user autonomy, raising questions about informed consent [52] and usability [24], [50]. Moreover, users may develop a *false sense of security* (aka over-trust) or become over-reliant on the companion [8], [65], [66]. As a middle path, Acquisti et al. [61] advocate for *soft paternalism* via nudges – tactics that steer behavior without completely restricting user choice. Non-blocking warnings that explain secure practices can be viewed as "soft paternalistic nudges." Hence, the companion should support, but not replace, the user agency by explicitly stating that the final decision remains with the user and by providing explanations that build the user's understanding over time. However, to effect meaningful change, users must also find the recommended behavior *feasible*, understand its *benefits*, and *agree* they want to adopt it [50]. Accordingly, xSec Companions should *simplify* secure tasks, *time* their prompts for moments when users are receptive, *justify* the potential benefits, and ultimately *respect* user autonomy throughout.

4) Fostering Trustworthiness & Minimize Privacy Risks: Trust remains a cornerstone of cybersecurity [67]. Users tend to prefer companions viewed as both *competent* and *empathic* [68], a combination that can be cultivated through reliable operation, meaningful feedback, and avoidance of invasive behaviors. As our workshop participants noted, users may feel uneasy if they perceive the companion as violating their privacy or operating beyond their control. Moreover, as our participants observed, the companion itself can become a *vector for privacy or security threats*. Sensitive data is frequently needed for context-aware functionality, and treating it carelessly could jeopardize user privacy. Consequently, xSec Companions should *minimize* the scope of data collection, *operate locally* whenever feasible (e.g., leveraging local large language models), and protect themselves from security threats. By maintaining robust safeguards and a transparent data policy, xSec Companions can also help ensure that users remain confident and comfortable in the ongoing interaction.

VIII. Conclusion

Traditional cybersecurity measures often struggle to instill long-term secure behaviors [24]. In contrast, *cybersecurity companions* offer a more engaging, user-centered approach by embedding assistance into daily routines through personalized interactions [6], [69], [70]. Yet, existing work offers limited guidance on how to design such companions effectively. To address this, we conducted ideation workshops with 12 participants to explore core design dimensions for cybersecurity companions. We synthesized these insights into the *xSec Companions Design Framework*, grounded in the Explainable Security (xSec) paradigm [11]. The framework provides a structured guide for researchers, developers, and designers. We also discuss key challenges – including warning fatigue, privacy, and autonomy – and highlight future opportunities such as AI-driven personalization and tangible, custom hardware. Our goal is to support the development of more dynamic, comprehensible, and effective cybersecurity companions.

Acknowledgment

We thank our study participants for their time and valuable feedback. This project has been funded by the European Union – NextGeneration EU and the dtec.bw – Center for Digitization and Technology Research of the Bundeswehr (projects MuQuaNet and Voice of Wisdom).

References

[1] S. Morgan, "Cybercrime to cost the world 10.5 trillion euro annually by 2025," *Cybercrime Magazine*, 2020.

[2] Verizon, "2021 Data Breach Investigations Report (DBIR)," https://www.verizon.com/business/resources/reports/2021/2021-data-breach-investigations-report.pdf, 2021.

[3] A. M. Alnajim, S. Habib, M. Islam, H. S. AlRawashdeh, and M. Wasim, "Exploring cybersecurity education and training techniques: A comprehensive review of traditional, virtual reality, and augmented reality approaches," *Symmetry*, vol. 15, no. 12, 2023. [Online]. Available: https://www.mdpi.com/2073-8994/15/12/2175

[4] J. Prümmer, T. van Steen, and B. van den Berg, "A systematic review of current cybersecurity training methods," *Computers & Security*, vol. 136, p. 103585, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404823004959

[5] J. Prümmer, "The role of cognition in developing successful cybersecurity training programs – passive vs. active engagement," in *Augmented Cognition*, D. D. Schmorrow and C. M. Fidopiastis, Eds. Cham: Springer Nature Switzerland, 2024, pp. 185–199.

[6] A. C. Tally, J. Abbott, A. M. Bochner, S. Das, and C. Nippert-Eng, "Tips, tricks, and training: Supporting anti-phishing awareness among mid-career office workers based on employees' current practices," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3544548.3580650

[7] Y.-M. Chiou, T. Barnes, C. Mouza, and C.-C. Shen, "Social robot teaches cybersecurity," in *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*, ser. IDC '20. New York, NY, USA: Association for Computing Machinery, Jul. 2020, pp. 199–204. [Online]. Available: https://dl.acm.org/doi/10.1145/3397617.3397824

[8] D. Pasquali, A. Kothig, A. M. Aroyo, J. E. Muñoz Cadorna, K. Dautenhahn, S. Bencetti, R. Francesco, and A. Sciutti, "That's not a good idea: A robot changes your behavior against social engineering," in *Proceedings of the 11th International Conference on Human-Agent Interaction*, ser. HAI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 63–71. [Online]. Available: https://doi.org/10.1145/3623809.3623879

[9] D. Benyon and O. Mival, "From human-computer interactions to human-companion relationships," in *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia*, ser. IITM '10. New York, NY, USA: Association for Computing Machinery, 2011, p. 1–9. [Online]. Available: https://doi.org/10.1145/1963564.1963565

[10] J. Zhang, N. M. Thalmann, and J. Zheng, "Combining memory and emotion with dialog on social companion: A review," in *Proceedings of the 29th International Conference on Computer Animation and Social Agents*, ser. CASA '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1–9. [Online]. Available: https://doi.org/10.1145/2915926.2915952

[11] L. Viganò and D. Magazzeni, "Explainable Security," in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, Sep. 2020, pp. 293–300. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9229719

[12] Y.-M. Chiou, T. Barnes, S. M. Jelenewicz, C. Mouza, and C.-C. Shen, "Teacher Views on Storytelling-based Cybersecurity Education with Social Robots," in *Proceedings of the 20th Annual ACM Interaction Design and Children Conference*, ser. IDC '21. New York, NY, USA: Association for Computing Machinery, Jun. 2021, pp. 508–512. [Online]. Available: https://dl.acm.org/doi/10.1145/3459990.3465199

[13] J. Guo, J. Guo, C. Yang, Y. Wu, and L. Sun, "Shing: A Conversational Agent to Alert Customers of Suspected Online-payment Fraud with Empathetical Communication Skills," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM, May 2021, pp. 1–11. [Online]. Available: https://dl.acm.org/doi/10.1145/3411764.3445129

[14] M. Pears, J. Henderson, and S. T. Konstantinidis, "Repurposing Case-Based Learning to a Conversational Agent for Healthcare Cybersecurity," in *Public Health and Informatics*. IOS Press, 2021, pp. 1066–1070. [Online]. Available: https://ebooks.iospress.nl/doi/10.3233/SHTI210348

[15] S. Adinolf, P. Wyeth, R. Brown, and R. Altizer, "Towards Designing Agent Based Virtual Reality Applications for Cybersecurity Training," in *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, ser. OzCHI '19. New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 452–456. [Online]. Available: https://dl.acm.org/doi/10.1145/3369457.3369515

[16] R. Minakawa and T. Takada, "Exploring alternative security warning dialog for attracting user attention: evaluation of "Kawaii" effect and its additional stimulus combination," in *Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services*, ser. iiWAS '17. New York, NY, USA: Association for Computing Machinery, Dec. 2017, pp. 582–586. [Online]. Available: https://dl.acm.org/doi/10.1145/3151759.3151846

[17] M. Silic, "Improving warning messages adherence: can Maya Security Bot advisor help?" *Security Journal*, vol. 33, no. 2, pp. 293–310, Jun. 2020. [Online]. Available: https://doi.org/10.1057/s41284-019-00185-7

[18] C. Bravo-Lillo, S. Komanduri, L. F. Cranor, R. W. Reeder, M. Sleeper, J. Downs, and S. Schechter, "Your attention please: designing security-decision uis to make genuine risks harder to ignore," in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, ser. SOUPS '13. New York, NY, USA: Association for Computing Machinery, 2013. [Online]. Available: https://doi.org/10.1145/2501604.2501610

[19] A. De Luca, B. Frauendienst, M. Maurer, and D. Hausen, "On the design of a "moody" keyboard," in *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, ser. DIS '10. New York, NY, USA: Association for Computing Machinery, Aug. 2010, pp. 236–239. [Online]. Available: https://dl.acm.org/doi/10.1145/1858171.1858213

[20] D. Napoli, S. Navas, S. Chiasson, and E. Stobert, "Something Doesn't Feel Right: Using Thermal Warnings to Improve User Security Awareness."

[21] Y. Do, L. T. Hoang, J. W. Park, G. D. Abowd, and S. Das, "Spidey Sense: Designing Wrist-Mounted Affective Haptics for Communicating Cybersecurity Warnings," in *Designing Interactive Systems Conference 2021*. Virtual Event USA: ACM, Jun. 2021, pp. 125–137. [Online]. Available: https://dl.acm.org/doi/10.1145/3461778.3462027

[22] Y. Wilks, S. Giles, and O. OX, "Artificial companions as a new kind of interface to the future internet," 01 2008.

[23] S. Danilava, S. Busemann, and C. Schommer, "Artificial conversational companions a requirements analysis," in *ICAART 2012*. SciTePress 2012, 2012.

[24] M. A. Sasse, J. Hielscher, J. Friedauer, and A. Buckmann, "Rebooting it security awareness – how organisations can encourage and sustain secure behaviours," in *Computer Security. ESORICS 2022 International Workshops*, S. Katsikas, F. Cuppens, C. Kalloniatis, J. Mylopoulos, F. Pallas, J. Pohle, M. A. Sasse, H. Abie, S. Ranise, L. Verderame, E. Cambiaso, J. Maestre Vidal, M. A. Sotelo Monge, M. Albanese, B. Katt, S. Pirbhulal, and A. Shukla, Eds. Cham: Springer International Publishing, 2023, pp. 248–265.

[25] S. Weber, M. Harbach, and M. Smith, "Participatory design for security-related user interfaces," *Proc. USEC*, vol. 15, 2015.

[26] I. Slesinger, L. Coles-Kemp, N. Panteli, and R. R. Hansen, "Designing through the stack: the case for a participatory digital security by design," in *Proceedings of the 2022 New Security Paradigms Workshop*, 2022, pp. 45–59.

[27] F. Quayyum, "Co-designing cybersecurity-related stories with children: Perceptions on cybersecurity risks and parental involvement," *Entertainment Computing*, vol. 52, p. 100753, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1875952124001216

[28] A. Hume, N. Ferreira, and L. Cernuzzi, "The design of a privacy dashboard for an academic environment based on participatory design," in *2021 XLVII Latin American Computing Conference (CLEI)*. IEEE, 2021, pp. 1–10.

[29] K. Andersen and R. Wakkary, "The magic machine workshops: Making personal design knowledge," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–13. [Online]. Available: https://doi.org/10.1145/3290605.3300342

[30] B. Rohrbach, "6–3–5 brainwriting," *bsatzwirtschaft*, 1969.

[31] hornetsecurity, "2024 cyber security report," last accessed on January 20, 2025. [Online]. Available: https://go.hornetsecurity.com/downloads/Cyber_Security_Report_2024_EN.pdf

[32] J.-Y. Jung, S. Qiu, A. Bozzon, and U. Gadiraju, "Great Chain of Agents: The Role of Metaphorical Representation of Agents in Conversational Crowdsourcing," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–22. [Online]. Available: https://dl.acm.org/doi/10.1145/3491102.3517653

[33] M. Wahde and M. Virgolin, "Conversational agents: Theory and applications," in *HANDBOOK ON COMPUTER LEARNING AND INTELLIGENCE: Volume 2: Deep Learning, Intelligent Control and Evolutionary Computation*. World Scientific, 2022, pp. 497–544.

[34] M. Koelle, K. Wolf, and S. Boll, "Beyond LED Status Lights - Design Requirements of Privacy Notices for Body-worn Cameras," in *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '18. New York, NY, USA: Association for Computing Machinery, Mar. 2018, pp. 177–187. [Online]. Available: https://dl.acm.org/doi/10.1145/3173225.3173234

[35] A. Rundi, "Brainwriting 6–3–5," in *The Innovation Tools Handbook, Volume 2*. Productivity Press, 2016, pp. 33–37.

[36] S. Delgado Rodriguez, M. Windl, F. Alt, and K. Marky, "The tapsi research framework - a systematization of knowledge on tangible privacy and security interfaces," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25. New York, NY, USA: Association for Computing Machinery, 2025. [Online]. Available: https://doi.org/10.1145/3706598.3713968

[37] M. Naeem, W. Ozuem, K. Howell, and S. Ranfagni, "A step-by-step process of thematic analysis to develop a conceptual model in qualitative research," *International Journal of Qualitative Methods*, vol. 22, p. 16094069231205789, 2023. [Online]. Available: https://doi.org/10.1177/16094069231205789

[38] S. T. Fife and J. D. Gossner, "Deductive qualitative analysis: Evaluating, expanding, and refining theory," *International Journal of Qualitative Methods*, vol. 23, p. 16094069241244856, 2024. [Online]. Available: https://doi.org/10.1177/16094069241244856

[39] V. Braun and V. C. and, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa

[40] V. Braun and V. Clarke, "Successful qualitative research: A practical guide for beginners," 2013.

[41] T. Gundu, "Chatbots: A framework for improving information security behaviours using chatgpt," in *Human Aspects of Information Security and Assurance*, S. Furnell and N. Clarke, Eds. Cham: Springer Nature Switzerland, 2023, pp. 418–431.

[42] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, ser. EASE '14. New York, NY, USA: Association for Computing Machinery, 2014. [Online]. Available: https://doi.org/10.1145/2601248.2601268

[43] J. Yoo and Y. Cho, "Icsa: Intelligent chatbot security assistant using text-cnn and multi-phase real-time defense against sns phishing attacks," *Expert Systems with Applications*, vol. 207, p. 117893, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422011435

[44] S. Lee, J. Lee, W. Lee, S. Lee, S. Kim, and E. T. Kim, "Design of integrated messenger anti-virus system using chatbot service," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2020, pp. 1613–1615.

[45] K. T. Reddy, "How deep learning chatbots empower cybersecurity against phishing attacks," n.a.

[46] F. Mosca and J. M. Such, "Elvira: An explainable agent for value and utility-driven multiuser privacy." in *AAMAS*, 2021, pp. 916–924.

[47] G. Ayci, A. Özgür, M. Şensoy, and P. Yolum, "Peak: Explainable privacy assistant through automated knowledge extraction," *arXiv preprint arXiv:2301.02079*, 2023.

[48] M. Nour, W. El Hefny, and A. El Bolock, "Cybot: A chatbot for teaching and testing cybersecurity courses," in *International Conference in Methodologies and intelligent Systems for Techhnology Enhanced Learning*. Springer, 2024, pp. 277–288.

[49] T. Wang, N. Zhou, and Z. Chen, "Cybermentor: Ai powered learning tool platform to address diverse student needs in cybersecurity education," *arXiv preprint arXiv:2501.09709*, 2025.

[50] J. Hielscher, A. Kluge, U. Menges, and M. A. Sasse, ""taking out the trash": Why security behavior change requires intentional forgetting," in *Proceedings of the 2021 New Security Paradigms Workshop*, ser. NSPW '21. New York, NY, USA: Association for Computing Machinery, 2022, p. 108–122. [Online]. Available: https://doi.org/10.1145/3498891.3498902

[51] A. Vance, D. Eargle, J. L. Jenkins, C. B. Kirwan, and B. B. Anderson, "The fog of warnings: how non-essential notifications blur with security warnings," in *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, 2019, pp. 407–420.

[52] P. Formosa, M. Wilson, and D. Richards, "A principlist framework for cybersecurity ethics," *Computers & Security*, vol. 109, p. 102382, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404821002066

[53] S. Delgado Rodriguez, A. Dao Phuong, F. Bumiller, L. Mecke, F. Dietz, F. Alt, and M. Hassib, "Padlock, the universal security symbol? - exploring symbols and metaphors for privacy and security," in *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 10–24. [Online]. Available: https://doi.org/10.1145/3626705.3627770

[54] E. Iturbe, E. Rios, A. Rego, and N. Toledo, "Artificial intelligence for next generation cybersecurity: The ai4cyber framework," in *Proceedings of the 18th International Conference on Availability, Reliability and Security*, ser. ARES '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3600160.3605051

[55] E. Casey and D. Chamberlain, "Capture the flag with chatgpt: Security testing with ai chatbots," in *19th International Conference on Cyber Warfare and Security: ICCWS*, 2024.

[56] N. Baym, L. Shifman, C. Persaud, and K. Wagman, "Intelligent failures: Clippy memes and the limits of digital assistants," *AoIR Selected Papers of Internet Research*, 2019.

[57] M. H. Bornstein, "Frames of mind: The theory of multiple intelligences," 1986.

[58] A. Adams and M. A. Sasse, "Users are not the enemy," *Commun. ACM*, vol. 42, no. 12, p. 40–46, dec 1999. [Online]. Available: https://doi.org/10.1145/322796.322806

[59] A. Whitten and J. D. Tygar, "Why johnny can't encrypt: A usability evaluation of PGP 5.0," in *8th USENIX Security Symposium (USENIX Security 99)*. Washington, D.C.: USENIX Association, Aug. 1999. [Online]. Available: https://www.usenix.org/conference/8th-usenix-security-symposium/why-johnny-cant-encrypt-usability-evaluation-pgp-50

[60] M. E. Zurko and R. T. Simon, "User-centered security," in *Proceedings of the 1996 workshop on New security paradigms*, 1996, pp. 27–33.

[61] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, Y. Wang, and S. Wilson, "Nudges for privacy and security: Understanding and assisting users' choices online," *ACM Comput. Surv.*, vol. 50, no. 3, Aug. 2017. [Online]. Available: https://doi.org/10.1145/3054926

[62] C. Nobles, "Stress, burnout, and security fatigue in cybersecurity: A human factors problem," *HOLISTICA–Journal of Business and Public Administration*, vol. 13, no. 1, pp. 49–72, 2022.

[63] N. Perry, M. Srivastava, D. Kumar, and D. Boneh, "Do users write more insecure code with ai assistants?" in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 2785–2799. [Online]. Available: https://doi.org/10.1145/3576915.3623157

[64] A. Habbal, M. K. Ali, and M. A. Abuzaraida, "Artificial intelligence trust, risk and security management (ai trism): Frameworks, applications, challenges and future research directions," *Expert Systems with Applications*, vol. 240, p. 122442, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423029445

[65] A. Haskard and D. Herath, "Secure robotics: Navigating challenges at the nexus of safety, trust, and cybersecurity in cyber-physical systems," *ACM Comput. Surv.*, vol. 57, no. 9, Apr. 2025. [Online]. Available: https://doi.org/10.1145/3723050

[66] J. Schoeffer, J. Jakubik, M. Vössing, N. Kühl, and G. Satzger, "Ai reliance and decision quality: Fundamentals, interdependence, and the effects of interventions," *J. Artif. Int. Res.*, vol. 82, p. 471–501, Apr. 2025. [Online]. Available: https://doi.org/10.1613/jair.1.15873

[67] D. Henshel, M. Cains, B. Hoffman, and T. Kelley, "Trust as a human factor in holistic cyber security risk assessment," *Procedia Manufacturing*, vol. 3, pp. 1117–1124, 2015, 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2351978915001870

[68] X. Cheng, X. Zhang, J. Cohen, and J. Mou, "Human vs. ai: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms," *Information Processing & Management*, vol. 59, no. 3, p. 102940, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457322000620

[69] D. Fanta, F. Sajid, M. Masssoth, and L. Kruck, "Efficiency of an artificial intelligence-based chatbot support for an it-awareness and cybersecurity learning platform."

[70] C. Kallonas, A. Piki, and E. Stavrou, "Empowering professionals: a generative ai approach to personalized cybersecurity learning," in *2024 IEEE global engineering education conference (EDUCON)*. IEEE, 2024, pp. 1–10.

[71] A. F. Westin, "Privacy and freedom," *Washington and Lee Law Review*, vol. 25, no. 1, p. 166, 1968.

[72] K. Salehzadeh Niksirat, D. Korka, H. Harkous, K. Huguenin, and M. Cherubini, "On the potential of mediation chatbots for mitigating multiparty privacy conflicts - a wizard-of-oz study," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CSCW1, Apr. 2023. [Online]. Available: https://doi.org/10.1145/3579618

[73] M. Simoni, A. Saracino, V. P., and M. Conti, "Morse: Bridging the gap in cybersecurity expertise with retrieval augmented generation," 2024. [Online]. Available: https://arxiv.org/abs/2407.15748

[74] P. Sahu, "How chatbot assistants can enhance social engineering attack detection?"

[75] ——, "Deep learning chatbot assistance for real-time phishing attack detection," n.a.

[76] ——, "Enhancing cybersecurity with 2fa and future chat-bot integration," n.a.

[77] A. Dan, S. Gupta, S. Rakshit, and S. Banerjee, "Toward an ai chatbot-driven advanced digital locker," in *Proceedings of International Ethical Hacking Conference 2018: eHaCON 2018, Kolkata, India*. Springer, 2019, pp. 37–46.

[78] F. Mosca and J. Such, "An explainable assistant for multiuser privacy," *Autonomous Agents and Multi-Agent Systems*, vol. 36, no. 1, Apr. 2022. [Online]. Available: https://doi.org/10.1007/s10458-021-09543-5

[79] V. Jüttner, M. Grimmer, and E. Buchmann, "Chatids: Explainable cybersecurity using generative ai," *arXiv preprint arXiv:2306.14504*, 2023.

[80] Y.-C. Fung and L.-K. Lee, "A chatbot for promoting cybersecurity awareness," in *Cyber Security, Privacy and Networking*, D. P. Agrawal, N. Nedjah, B. B. Gupta, and G. Martinez Perez, Eds. Singapore: Springer Nature Singapore, 2022, pp. 379–387.

[81] M. Kaheh, D. K. Kholgh, and P. Kostakos, "Cyber sentinel: Exploring conversational agents in streamlining security tasks with gpt-4," *arXiv preprint arXiv:2309.16422*, 2023.

[82] V. Jüttner, M. Grimmer, and E. Buchmann, "Chatids: Advancing explainable cybersecurity using generative ai," *International Journal On Advances in Security*, vol. 17, no. 1, p. 2, 2024.

[83] M. Hoffmann and E. Buchmann, "Chatsec - towards enhancing security vulnerability reports for non-experts," Mensch und Computer 2024 - Workshopband, 2024.

[84] G. Ayci, A. Özgür, M. Sensoy, and P. Yolum, "Explain to me: Towards understanding privacy decisions," in *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, 2023, pp. 2790–2791.

[85] F. Mosca, J. Such, and P. McBurney, "Towards a value-driven explainable agent for collective privacy," in *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems*, 2020.

[86] A. C. Kurtan and P. Yolum, "Assisting humans in privacy management: an agent-based approach," *Autonomous Agents and Multi-Agent Systems*, vol. 35, no. 1, p. 7, 2021.

[87] G. Misra and J. M. Such, "Pacman: Personal agent for access control in social media," *IEEE Internet Computing*, vol. 21, no. 6, pp. 18–26, 2017.

[88] ——, "React: Recommending access control decisions to social media users," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 421–426.

[89] R.-D. RAMON, J. Alemany, H. STELLA, and A. Garcia-Fornes, "Automatic generation of explanations to prevent privacy violations," 2019.

[90] R. L. Fogues, P. K. Murukannaiah, J. M. Such, and M. P. Singh, "Sosharp: Recommending sharing policies in multiuser privacy scenarios," *IEEE Internet Computing*, vol. 21, no. 6, pp. 28–36, 2017.

[91] C. Paduraru, C. C. Patilea, and A. Stefanescu, "Cyberguardian: An interactive assistant for cybersecurity specialists using large language models," *Proc. of ICSOFT*, vol. 24, pp. 442–449, 2024.

[92] B. Filar, R. Seymour, and M. Park, "Ask me anything: A conversational interface to augment information security workers." in *SOUPS*, 2017.

[93] I. Gulenko, "Chatbot for it security training: Using motivational interviewing to improve security behaviour." in *AIST (supplement)*, 2014, pp. 7–16.

[94] S. Kowalski, K. Pavlovska, and M. Goldstein, "Two case studies in using chatbots for security training," in *Information Assurance and Security Education and Training: 8th IFIP WG 11.8 World Conference on Information Security Education, WISE 8, Auckland, New Zealand, July 8-10, 2013, Proceedings, WISE 7, Lucerne Switzerland, June 9-10, 2011, and WISE 6, Bento Gonçalves, RS, Brazil, July 27-31, 2009, Revised Selected Papers 8*. Springer, 2013, pp. 265–272.

APPENDIX

A. Study Guide for the Introduction Phase

The following Sections describe how we introduced the study's topic to our participants.

1) Explanation of Phishing: A Phishing attack has the underlying goal of acquiring sensitive information, such as passwords or bank account details, from the victim. Phishing attacks often happen through e-mails or take place on websites. The attacker masquerades as a legitimate business or person. The attackers often create a sense of urgency and tend to threaten the victim.

On this slide, you can see an example with tips on how to detect a phishing email. The words "important update" marked as number 1 create a sense of urgency, because the victim should believe that there's a really important update that needs to be installed immediately. Number 3 shows that, when hovering over the written link with the mouse cursor, a suspicious-looking URL appears. The steps at number 2 show a strange request to click on a link for a website, enter account details, and select "service update" to receive an update. The attacker needs the user to perform an action in order for the Phishing attack to be successful. As seen at number 4, the attackers ask for sensitive information, in this case, account details, consisting of username and password, to be able to receive the update. Such sensitive information is something that a legitimate business would not need to know for an update and would never ask for. In addition to that, Phishing emails often have mistakes in spelling or writing. And they can also address the victim with the wrong name or have no personalization at all.

2) Explanation of Privacy: "Privacy is the right to prevent the disclosure of personal information to others" [71]

3) Explanation of IoT: IOT stands for Internet of Things. IoT includes intelligent devices, networks, and systems. Smart homes are IOT devices that are installed in a home. To make it clearer what this means exactly, I'll show you some examples. The first picture shows a doorbell camera from the company "Ring", which shows the person inside the house who is standing in front of their door. In the second picture, you can see a rather well-known example, an Amazon Echo Dot, a virtual assistant you can speak with. The last picture on the slide shows an electronic door lock, where you can use your mobile phone to unlock doors. Privacy in IOT is all about whether a person agreed to their data being recorded. For example, if Alexa is still listening to its surroundings even though it's not supposed to. Or when a doorbell camera records a postman without his consent. We see that there are some critical problems that can occur in the field of cybersecurity, as we saw with the Phishing attacks and the topic of privacy in IOT.

4) Explanation of Social Companions: Now I'd like to introduce Social Companions because they could help us when it comes to cybersecurity. A Social companion can help us with cybersecurity in many different ways. For example, they could act as a warning system to warn us, for example, of malicious links or websites. They could also teach about cybersecurity and spread awareness. Social Companions can have conversations with the User through speaking with them, through using their emotions to give feedback or writing messages like a chatbot and much more.
A social companion can be many different things. On the slide, we can see various ways that a Social Companion can look. It can look more or less human. From left to right, we can see that Social Companions can look like a god-like entity. They can also look like a human, an animal, or a plant. Last but not least, Social Companions can also look like objects. It's hard to imagine what a god-like entity looks like, so I brought you some well-known examples that you might be familiar with.

[...] [Showing and explaining the example companions listed in C one by one]

As we saw before, a Social Companion can be physical, meaning a tangible object that one can touch, like the privacy flower, the Tamagotchi, or the robot. However, they can also be virtual, like the Security Warning Popup, the virtual assistants, or the CiSA chatbot.

B. Task Instructions

The following Sections describe how we explained the study tasks to our participants.

1) Brainwriting: Now that you know a bit about the cybersecurity topics that we will focus on today and what a social companion is, I'd like to do something called brainwriting with you. Brainwriting is similar to brainstorming. It's a method to gather ideas for social companions for cybersecurity. These ideas can also serve as an inspiration for the prototypes that we'll create afterwards. So we will now split into groups of 2 (or 3) people. [...] I've prepared some papers for that. Every group member will get a paper. On these papers are questions about Social Companions. On one paper are two questions for the cybersecurity Topic of Phishing, and on the other paper are the same two questions for Privacy in IoT. The idea is that you answer just question number one on the paper. After that, you will exchange the paper with your group member. Then you'll read what your group member wrote for the first question on the new paper you got and answer the second question. This means you'll add onto the idea that's already written on the paper, and you can also add your own ideas to it.

2) Crafting Session: Each group will now design a prototype for a Social companion for one of the two cybersecurity topics I presented to you before, with the help of the ideas and inspiration that we gathered through the brainwriting that we just did. If you have too many different ideas to just put them into one Companion, or if you just want to create an additional one that looks different or functions differently, you are also allowed to create multiple prototypes. So the plan is that each group creates at least one prototype. Half of the groups will create prototypes for Phishing, and the other half of the groups for Privacy in IoT. [...] You can use all the materials you see here. How you'll design your prototype is completely up to your imagination. You can design your companion however you want. You should also think of a

name for your companion. Please write the name of your Companion and your group number onto a post-it note and stick it somewhere onto your prototype.

C. Image Resources for Example Companions

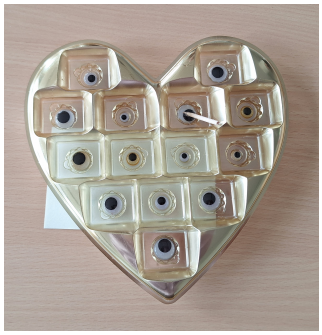
- Animated Dog Avatar: Figure 1 from [16]
- Bixby: Image from https://store-images.s-microsoft.com/image/apps.50257.14239178866294621.2f72ce3e-9285-48ba-99c2-9d22d30d507a.86b56be3-e259-4e21-a765-4b1efb23f045?h=464, last accessed November 2, 2023
- (CiSA) Chatbot: Image from https://www.augsburg.de/fileadmin/_processed_/6/f/csm_230728_CISA_Chatbot_Logo_Rund_910aa429c3.jpg, last accessed November 2, 2023
- Cortana: Image from https://news.microsoft.com/de-de/cortana-oeffnet-sich-neue-entwicklerwerkzeuge-erweitern-zugang-zur-digitalen-assistentin/, last accessed November 2, 2023
- Human-like Agents: Figure 1 from [33]
- Privacy Flower: Right part of Figure 2 from [34]
- Siri: Image from https://www.apple.com/v/siri/h/images/overview/workout_tile_2__fduarvg890qe_medium_2x.png, last accessed January 1, 2025
- Tamagotchi: https://shop.bandai.co.uk/wp-content/uploads/2024/08/bandai-tamagotchi-tamagotchi-original-rainbow7.jpg, last accessed January 1, 2025
- Zenbo Social Robot: Figure 1 from in [7]

E. Snowball-Sampling based Literature Search

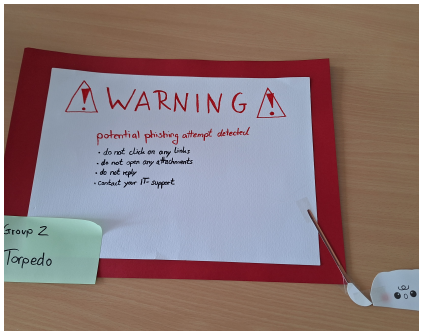
TABLE IV: Publications found through a three-round snowball sampling approach [42] starting from 8 publications (see round 0). The "cited by" publications were revised on January 16, 2025.

round	publication	references	cited by	newly found
0	Towards Designing Agent Based Virtual Reality Applications for Cybersecurity Training [15]	25	14	-
0	Social Robot Teaches Cybersecurity [7]	15	5	[12]
0	Shing: A Conversational Agent to Alert Customers of Suspected Online-payment Fraud with Empathetical Communication Skills [13]	52	15	[72]
0	That's not a Good Idea: A Robot Changes Your Behavior Against Social Engineering [8]	62	1	-
0	Repurposing Case-Based Learning to a Conversational Agent for Healthcare Cybersecurity [14]	17	17	-
0	Improving warning messages adherence: can Maya Security Bot Advisor Help? [17]	55	7	-
0	Exploring Alternative Security Warning Dialog for Attracting User Attention: Evaluation of "Kawaii" Effect and Its Additional Stimulus Combination [16]	14	9	-
0	ICSA: Intelligent Chatbot Security Assistant Using Text-CNN and Multi-Phase Real-Time Defense Against SNS Phishing Attacks [43]	37	23	[41], [44], [73]–[77]
1	Teacher Views on Storytelling-based Cybersecurity Education with Social Robots [12]	18	10	-
1	Design of Integrated Messenger Anti-Virus System Using Chatbot Service [44]	15	6	-
1	On the Potential of Mediation Chatbots for Mitigating Multiparty Privacy Conflicts - A Wizard-of-Oz Study [72]	133	11	[78]
1	Toward an AI Chatbot-Driven Advanced Digital Locker [77]	12	9	-
1	Chatbots: A Framework for Improving Information Security Behaviours using ChatGPT [41]	39	21	[45]
1	MoRSE: Bridging the Gap in Cybersecurity Expertise with Retrieval Augmented Generation [73]	96	2	[49], [55], [79]
1	How Chatbot Assistants Can Enhance Social Engineering Attack Detection? [74]	6	2	-
1	Enhancing Cybersecurity with 2FA and Future Chat-bot Integration [76]	13	0	[80]
1	Deep Learning Chatbot Assistance for Real-Time Phishing Attack Detection [75]	9	0	-
2	ChatIDS: Explainable Cybersecurity Using Generative AI [79]	29	19	[81]–[83]
2	Capture the Flag with ChatGPT: Security Testing with AI ChatBots [55]	2	4	-
2	An Explainable Assistant for Multiuser Privacy [78]	95	27	[46], [47], [84]–[90]
2	How Deep Learning Chatbots Empower Cybersecurity Against Phishing Attacks [45]	9	0	-
2	CyberMentor: AI Powered Learning Tool Platform to Address Diverse Student Needs in Cybersecurity Education [49]	36	0	[70]
2	A Chatbot for Promoting Cybersecurity Awareness [80]	15	16	[48], [69], [91]–[94]
3	Cyber Sentinel: Exploring Conversational Agents in Streamlining Security Tasks with GPT-4 [81]		– not screened–	
3	ChatIDS: Advancing Explainable Cybersecurity Using Generative AI [82]		– not screened–	
3	ChatSEC - Towards Enhancing Security Vulnerability Reports for Non-Experts [83]		– not screened–	
3	ELVIRA: An Explainable Agent for Value and Utility-Driven Multiuser Privacy [46]		– not screened–	
3	Towards a value-driven explainable agent for collective privacy [85]		– not screened–	
3	PEAK: Explainable Privacy Assistant through Automated Knowledge Extraction [47]		– not screened–	
3	Assisting Humans in Privacy Management: An Agent-Based Approach [86]		– not screened–	
3	Pacman: Personal Agent for Access Control in Social Media [87]		– not screened–	
3	Automatic Generation of Explanations to Prevent Privacy Violations [89]		– not screened–	
3	Empowering Professionals: A Generative AI Approach to Personalized Cybersecurity Learning [70]		– not screened–	
3	CyberGuardian: An Interactive Assistant for Cybersecurity Specialists Using Large Language Models [91]		– not screened–	
3	Cybot: A Chatbot for Teaching and Testing Cybersecurity Courses [48]		– not screened–	
3	Efficiency of an Artificial Intelligence-based Chatbot Support for an IT-Awareness and Cybersecurity Learning Platform [69]		– not screened–	
3	Ask Me Anything: A Conversational Interface to Augment Information Security Workers [92]		– not screened–	
3	Chatbot for IT Security Training: Using Motivational Interviewing to Improve Security Behaviour [93]		– not screened–	
3	Two case studies in using chatbots for security training [94]		– not screened–	

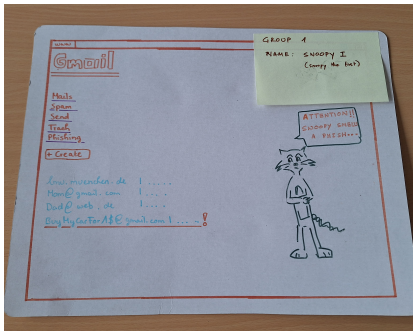
D. Photos of Participant's Crafted Prototypes



(a) G1 - Privacy in IoT: Orwelly

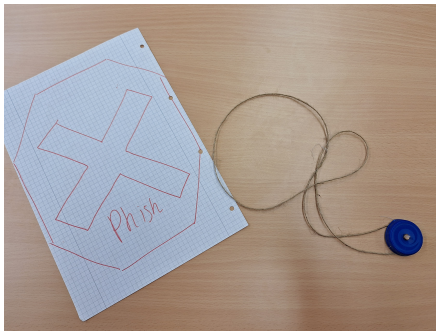


(b) G2 - Phishing: Torpedo

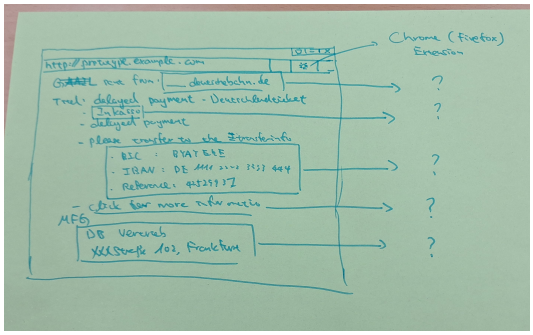


(c) G3 - Phishing: Snoopy I

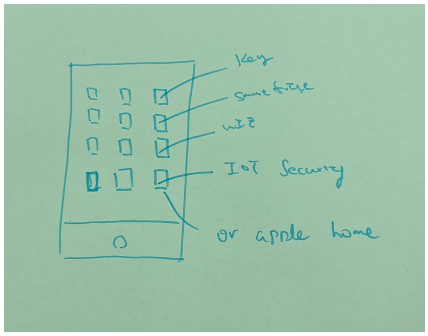
(d) G4 - Privacy in IoT: Proofy01



(e) G5 - Phishing: Safory



(f) G5 - Phishing: Catch It



(g) G5 - Privacy in IoT: Secure+

Fig. 3: The participants (G1-G5) of our three ideation workshops crafted seven different paper prototypes of cybersecurity companions for either phishing or privacy in IoT.