User-Agentified Security

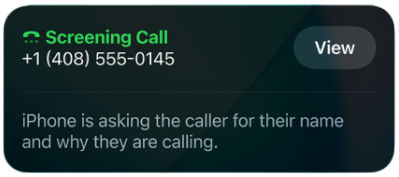
Filipo Sharevski
DePaul University

Abstract

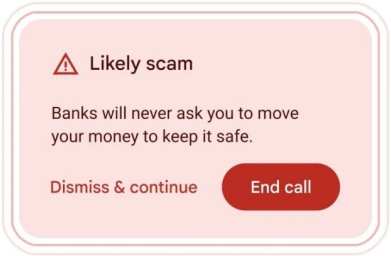
The term *user-agentified security* is introduced in this paper to refer to users' security decision making involving an agentic facilitation in some form. User-centered security has helped so far in this capacity, though always leaving the final decision entirely to the end users – (master) passwords are still out choice and we still decide on what links to click or unsolicited phone calls to pick. Agentic AI reasoning now offers a tempting alternative to seize the final decision from users and forever rid of weak passwords, stolen credentials, or empty bank accounts. Should this alternative materialize, users will inevitably move from the central position in security decision making to consolingly occupy a place "in the loop." But what it means to be in the loop with an AI agent when it comes to security? This paper proposes a research agenda to capture, define, and elucidate how users, but AI agents too, conceptualize user-agentified security. It does also reflects on the mutual trust as a factor sustaining any combination of roles, contexts, and tasks taken by users and AI agents in a security decision making relationship. Doing so, touches upon the facets of rules, ethics, and societal forces that inevitably shape the transition from user-centered to user-agentified security.

1 Introduction

If you have a recently updated iOS or an Android smartphone, you have probably experienced a new "*call screening*" feature (see Figure 1). The phone itself picks up the call, asks the caller a reason for calling and to state your name, and in the meanwhile, on the locked screen shows a live transcript of the call, in case you would like to pick up (alternatively, a warning summarizing the exact call content). This is user-centered security at its best: frictions, interventions, and warnings, all with the goal to offer as much a possible information and buy as much as feasible time for a user to fend of a potential scam call. The final decision is still in the hand of the users if they want to pick up, leave it ring, or let it go to a voice mail (it could be the doctor's office, who knows [28]).



(a) iOS



(b) Android

Figure 1: Call Screening Feature

This is not a fully fledged AI agent, yet, but more so an AI-enabled assistant, one would object. All the same, it does serve an illustrative purpose to imagine a shift from a "*call screening*" assistant to an autonomous AI agent that handles all calls, scam included, for users. No transcripts, no warnings, no frictions, but simply a report, preferably once a month over email, on the number and types of scam calls fended off, with a brief summary of their content. Better, this AI agent – let's call it *Wintermute* [9] – doesn't do the job just for voice calls, but video too (FaceTime, Zoom, etc); fends off face swap

Workshop on Designing Human–AI Teaming for Security (HATS), co-located with the Symposium on Usable Privacy and Security (SOUPS) 2026.
August 23, 2026, Hannover, Germany.

deepfake calls from what appears to be a scammer posing as a family member, urgent conferences calls with your boss calling on supposedly their private number, and so on. Now, the final decision is actually in the hand of the AI agent – not the users – and *Wintermute* decides who deserves the honor of picking up, who should be treated with a 'leave it ring,' or screened through a voice mail, for a change.

What these two paragraphs captured is the slow, but sure transition from user-*centered* to user-*agentified* security. Employing usability to minimize the negative outcomes of user decisions on the overall security of systems [40] has been a largely successful effort [7], but there is no guarantee that usability alone could do the job going ahead. The reason is not that users have a notoriously dreadful track record of poor security judgments and will keep doing so in the face of what I call "*usAIble security*" (the use of AI for usable security purposes; an example is the aforementioned *call screening* intervention), but more so the rise of what I see as "*artiphishial intelligence*" (the use of any AI for deception purposes; deepfake calls, cloned voices, etc. [29]). Making a room for AI agentification, in addition to usability, does however affects the users' security decision making balance – so far, as sole decision-makers they occupied the center (or the users made the final decision), but now there are *two* decision makers, so the center no longer makes a space for both (where does the 'buck stop' for the final decision is an open question).

The goal of this paper, thus, is to proactively pave the way towards *user-agentified* security, discussing the concepts, relationships, ethics, and open questions shaping the renegotiation of the users' position in security decision making. My position, openly and deliberately, is still 'user first' despite the tempting pull of 'AI first' – the shorthand for keeping the users "in the loop." The notion of keeping users "in-the-loop" sets any agentic reasoning explicitly in the center of security in the sense that the final decision is almost always theirs, with the users' "involvement" *passive* and bound to be eventually phased out. On the contrary, the "user-agentified security" sees the users' involvement in an *active* regard and they are here to stay (if not for altruistic, then for liability reasons [1]).

User-centered security from the very beginning threw its weight in designs that fit the users' (incorrect) "*mental models*" or what they conceived of security while balancing against their poor judgments [4]. Introducing AI agents that could not just correctly "conceive" of security, but always make perfectly right judgments turns the "*one dimensional*" problem of fitting the mental models to a "*two dimensional*" one of role delegation, conflict resolution, improvement, and cooperation. The "in-the-loop" approach eschews the second dimension and simply transfers the one dimensional problem from *mental* to *AI* modeling. In contrast, the user-agentified security cares to develop "*joint mental models*" or "*user/agentic mental models*." We might know what users conceive of security, but we are yet to learn what users conceive of getting to do security *with* AI agents. An equally unchartered territory is what AI agents conceive of users in addition to perfect system security. If a loop is established, none of this knowledge would be able to produced, refined, and ultimately used towards internalizing a proper *secure behavior*. And we care for secure behavior because it casts "security as being" rather than "security as survival" [20] in a life worth living for users in times of an increasingly AI-constructed reality.

[1] An argument that additionally challenges the reductive "in-the-loop" framing comes from the way security is embedded in the order of liberal-democratic societies. Security has never been agreed upon as a public good and therefore bears a private good, capitalistic connotation, i.e., *security is not free* (and someone has to pay for it) [30]. If the "loop" allows security to be payed for/to the maker of the AI agent replacing the humans as final decision makers, then any exploit in this protection arrangement would set the maker liable for the security failure. Makers might pay, but its hard to believe that agentic security enterprises (read BigTech) would do [11]. Users, as always, have to stay to pick up the tab. Otherwise, who doesn't want to replace the whole security decision making palaver with a $9.99/month "out-of-loop" *Wintermute+* AI agent that does the anti scam/phishing on our behalf?

2 Mental Models: *It Takes Two to Tango*

User-*agentified* security understands (at least) a two-dimensional space of mental models between users and agents, some of which are friends, and some of which are foes, as shown in Table 1. Humans – abusers, but system designers too – always thought of ordinary users as the "weakest link in security" from the get-go, either that they are exploitable or they need security to be usable to avoid exploitation [24]. We are yet to find out what humans think of "usAIble security agents" though work has been done in the realm of human-agent teaming that could inform at least the dispositions between authentic and synthetic personas in collaborative contexts [26]. While AI agents tend to be conceptualized as an ideal doppelgänger version of users [39], security is a real test whether the strength lies in a user-agentic collaboration. It is especially important to keep in mind that users would, outside of security, increasingly see AI as a cohabitant in various forms of (anthropomorphic) AI clones, so the 'benefits' of user-agentified security might come at a cost/threat of identity fragmentation, agentic phobia, over-attachments [16], but also parasociality [18]. A helpful way ahead would be to elicit these models both in the context of the 'inner functioning' and the 'operations' of "usAIble security agents" [25].

Table 1: User-Agentified Mental Models of Security

		Agents	
	User	**UsAIble**	**Artiphishal**
User	the "weakest link"?	human-agent teaming, doppelgängers	synthetic content, avatars, personas
UsAIble	a *human* friend to aid?	an *agentic* ally?	an *agentic* foe?
Artiphishal	a human foe to exploit?	an *agentic* foe?	an *agentic* ally?

What users thought of security or the risks associated with abusing systems was not always the correct representation of how security [4, 36] and the involved technology [33] actually works. There were always the concepts of physical, medical, criminal, military, or market models, but these were implicitly connected to other humans – abusers – that have the intention to compromized the security, often through exploiting ordinary users. The "artiphishialy intelligent agents" put an extra premium on deception, both in context of varieties, but also interaction. What this means is that we are yet to see whether and how users conceptualize the use of AI for deception – deepfake content *and* deepfake personas (represented through malicious AI avatars) [12]. Evidence is piling up about the users' inability to spot a deepfake [3, 35], but we yet to gather an idea of what users make of using these deepfakes in context of phishing or scamming [29]. For some time, users have been interacting with AI generated avatars or personas, for example on social media [19] or through Siri, Alexa, and the likes [27]. But here one is pressed to think about the displacement of trust and intentionality between cases where synthetic personas make unintentional mistakes [13] an when these are deliberately made for deception purposes [2].

So far, this is still a very user-centric thinking. What really is the task ahead of user-*agentified* security is to think of what the *agents*, both friends and foes, would conceive of the users and their counterparts in (ab)using security of systems. Are the "usAIbly" friendly ones going to be designed to comply, conflict, challenge, or overrule humans [1] assuming they don't hold them in a high (security) regard? Or the opposite, they would be there to aid (in kind of a 'second opinion' style) the hereto chronically poor security judgments [17]? How much should they help before this aid becomes counter-productive and the promise of *extended cognition* becomes not a security, but a problem that manifests in the *users' cognition* elsewhere, from productivity to interpersonal relationships [23]? Equally, there is the collaboration aspect of teaming with other "usAIble security agents" (e.g., *Wintermute* handles my calls, but would it be good to talk to its email *Neuromancer* counterpart to spot a coordinated phish/scam campaign against me?). An allyship seems an obvious choice, but AI agent providers might not be that cozy in what would certainly appear as a competitive market for one-stop-shop of agentic anti-deception services [8].

And then we have the "artiphishally intelligent" yet malicious agents to reckon with. Threat modeling was always the name of the game in user-centered security, but here we are at task to '*model the model*' of the threat. Users are here to be deceived and that might seem increasingly trivial for malicious AI agents (content, persona, avatars), but what about meeting their *own* kind? Deception and manipulation are in the eye of the beholder, and we always thought of them in the context of violating the basic principles of security, but would these 'rules' be observed in an AI-vs-AI security game [31]? Users, by the standard account of phishing and scamming, were led *away* from beliefs and behaviors about systems. Would this means that a successful deception by an "artiphishally intelligent" agent would not just violate the security but introduce doubt, uncertainty, and distrust in the relationship between users and AI agents? And if so, would they find "partners in crime" with other "artiphishally intelligent" agents to break the user-agentified protection relationship?

3 Reflections on Trusting Trust: *Who AI You?*

The two-dimensionality of the decision-making space naturally demands user-*agentified* security to reinterpret the trust disposition of both users and agents in similar regards. Users, making final security decisions, had little reason not to trust user-centered interventions (nudges, frictions, warnings, etc.) as they were deliberately designed not to interfere with the users' freedom and final judgment. Leaving the realm of *subtle influence* towards a full blown *decision participation*, Ken Thompson's rhetorical question places an important premium on trusting agents, or better, "on the people who wrote the software that then wrote the agentic security software" [32].

Should the users now, if they want to remain involved in the security decision making, turn towards a Reganesuque '*trust, but verify*' outlook towards agents, for starters, and then to a full blown '*zero trust*' security disposition (in principle, not just about authentication)? It seems that such an approach would be rather burdensome in chasing the possibility of a '*rogue*' AI agents, and besides, it would be a real challenge for users to evaluate whether a security judgment was a matter of false positive/negative classification or a deliberate decision to undermine security. Again, it is tempting to turn to the "loop" as a way of offloading this burden from users, though this puts an unrealistic expectation that the AI agents will always get security right. As things stand, users might be 'fooled once' if their security AI guardian fails (shame on them), but hardly 'fooled twice' as then trust is rather lost (shame on us).

This highlights the critical role usability played vis-à-vis security as it brought both convenience but also established trust in usable security interventions. If users have to replace their final decision making role with a role of verification and audit of AI agents, then the benefit of supposedly improved security decision making is not worth neither the cost of motivation nor the investment in doing user-*agentified* security right in the first place. Historically, though, user-centered security has also been into a position to *earn* the trust of users, both because they felt they will be sold short on the security mechanisms' correctness side and on the convenience side too [37]. The long road to trusting user-centered interventions is paved with failures, such as the attempt to ensure correctness (or "*openness*") of PGP for user email encryption. While it came to a too high of a convenience cost (despite its correctness), rendering it pretty much unusable [38], it did show that allowing for *redundancy* (*reproducibility*), that is, being *open* could earn the hearts and minds of users.

Users don't have to verify every agentic decision and a suspected outcome could be traced back to at least a 'bad faith' agentic release, akin to Ubuntu and Linux distribution openness, for example. This does, however, indicates that user-*agentified* security begets a broader community or users to succeed, something that is in tensions with the hereto individualistic (and market, of course) nature of the user-centered security. Not to mistake the forest for the trees, the goal, at least for now, is to demonstrate that agents are trustful decision makers shoulder-to-shoulder with users. So the idea of openness is at least worthy of considering *open-source agents* for user-*agentified* security as a good test of the whole idea from moving away from the center, at least for now. AI does have roots in *open source* so there is a basis for trying so [14].

4 Social Order: *Green Dollars and Red Tape*

Enthusiastic as the *open source* road might be, when it comes to AI does always leads to closed gates, as a the societal order makes no room for free infrastructure, storage, training and so. It would be nice to see commodified "usAIble security agents," but a more realistic rendition of the user-agentified dynamics is an oligopolistic, winner-takes-all outcome [21] where the makers that provide the current user-centered security would do so in the next iteration (plus, obviously, some new "disruptors"). And where there is (security) BigTech, there is the attention of the regulators, advocates (privacy, yes), policy makers, national security planners, among many.

So far, the tensions revolved around "human (more fittingly, consumer) rights" but now we have two entities that render security decisions that are in position to demand rights. First question is, who is going to be liable when a phishing email or a scam call actually results in a breach of a company? Next, one natural wonders what would happen with the consumer protections. Would FTC, FCC, and the likes of any 'commission' be keeping an eye on the users or now they have agents to represent and cater to too? With standardization, then, one also reckons whether we would see a NIST "special publication" dedicated to AI agentic security only[2], or a full gamut of relationships, roles, conflicts and conflict resolution, and *user*-agentic cooperation policies that involve the users too? Critical infrastructure also comes to mind and CISA, so far, has issued an advisory on a 'careful adoption' but this tell us little about what and how the user-agetified life looks a like when protecting the national security [3]

User-*agentified* privacy surely deserves an entire treatise in equal measures as one thing was, in the past, to get privacy breached, infringed, and contravened by an *external* identity, context or no context [22], but an entire different game is when this happens by an *internal* one, in a form of an AI agent. As this is a two-way street, regulation would be in the hot seat facing pressure from the neoliberal order that seek AI agents in the security decision making process for profit purposes. Objections to security and privacy regulation would inevitably crop up [11] to define what the users (read markets) really like and expect from the agentificaiton of the user-centered protections. User-*agentified* security implementations would, by the virtue of agentic participation, would push regulations to consider removing barriers to entry or effectively endorsing the right of an AI agent to participate in a security decision making from the get-go with a well thought-out playbook, but then, it would have to balance out for the user rights that might eventually clash with this endorsement. If this is too convoluted, regulation might even push for incentives where the user involvement is eschewed in favor of full a agentic-to-agentic, *post-user* security altogether.

Investment in cybersecurity is what shapes even the current usability offerings for users so one could expect for this to be the case too. The cautionary tale of passwords giving way to passkeys [15], in its core, its not just about the importance of usability or correctness but more so about the fact that even security is touched by the '*invisible hand*' of the markets. A reflection back on the (regulatory) evolution on what constitutes a 'strong password' and strict 'password composition policies' [10] shows that usability was never a bet for investment [6] and instead market entrepreneurs have put their money where the 'hack' is, i.e., intrusion detection, monitoring, DevOps, and such. What this tells about user-*agentified* security is that the agentic evolution might impose *opportunity costs* on a *post-user security* where agents need not to bother with users and all. Even if regulation deems user-*agentified* security as a 'safe harbor' for liabilities, it might well reflect, as past have shown, an artificial (what an linguistic irony) rather than a natural scarcity of security.

5 Ethics: *The User is Dead, Long Live the User!*

The case for keeping the users' hold on the final security decision in the treatise of the user-agentified security was driven towards the preservation of *moral autonomy* in the face of what might eventually morph into an 'algorithmic autonomy' [34]. An agentic help would inevitably mimic an 'exert' version of the user and doing so runs the risk of undermining the users' autonomy but also dignity to make mistakes, even security ones, *on their own terms* [5]. Users' liberty and anonymity are at stakes too, as "usAIble security agents," despite the best of intentions, would be provisioned by private market entities that would be tempted (or simply compelled by governments) to glean into the personal information of users for the sake of security. Fairness as a value does require a thorough look at the distributional effects of the new decision making partnership (bias, of course). If users always pick up the tab of security failures, then so much for the justice legitimatizing user-*agentified* security as a worthy successor of the user-*centered* one.

[2] Open for RFCs: AI Agent Standards Initiative
[3] Careful Adoption of Agentic AI services

References

[1] David Almog, Romain Gauriot, Lionel Page, and Daniel Martin. AI Oversight and Human Mistakes: Evidence from Centre Court. In *Proceedings of the 25th ACM Conference on Economics and Computation*, EC '24, page 103–105, New York, NY, USA, 2024. Association for Computing Machinery.

[2] Nikola Banovic, Zhuoran Yang, Aditya Ramesh, and Alice Liu. Being trustworthy is not enough: How untrustworthy artificial intelligence (ai) can deceive the end-users and gain their trust. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW1), April 2023.

[3] Efe Bozkir, Clara Riedmiller, Athanassios N. Skodras, Gjergji Kasneci, and Enkelejda Kasneci. Can you tell real from fake face images? perception of computer-generated faces by humans. *ACM Trans. Appl. Percept.*, 22(2), November 2024.

[4] L. Jean Camp. Mental models of privacy and security. *IEEE Technology and Society Magazine*, 28(3):37–46, 2009.

[5] Eleonora Catena. Ai and human autonomy: a literature review. *AI and Ethics*, 6(1):126, 2026.

[6] Dinei Florêncio and Cormac Herley. Where do security policies come from? In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, New York, NY, USA, 2010. Association for Computing Machinery.

[7] Anjuli Franz, Verena Zimmermann, Gregor Albrecht, Katrin Hartwig, Christian Reuter, Alexander Benlian, and Joachim Vogt. SoK: Still plenty of phish in the sea — a taxonomy of User-Oriented phishing interventions and avenues for future research. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 339–358. USENIX Association, August 2021.

[8] Dan Geer, Eric Jardine, and Eireann Leverett. On market concentration and cybersecurity risk. *Journal of Cyber Policy*, 5(1):9–29, 01 2020.

[9] William Gibson's. *Neuromancer*. Springer, 1984.

[10] Paul A. Grassi, James L. Fenton, Elaine M. Newton, Ray A. Perlner, Andrew R. Regenscheid, William E. Burr, Justin P. Richer, Yee-Yin Choong Kristen, K. Greene, and Mary F. Theofanos. Nist special publication 800-63b, 2026.

[11] Mark W. Hodgins. The perils of cybersecurity regulation. *The Review of Austrian Economics*, 38(4):391–411, 2025.

[12] Ilkka Kaate, Joni Salminen, João M. Santos, Soon-Gyo Jung, Hind Almerekhi, and Bernard J. Jansen. "there is something rotten in denmark": Investigating the deepfake persona perceptions and their implications for human-centered ai. *Computers in Human Behavior: Artificial Humans*, 2(1):100031, 2024.

[13] Antino Kim, Mochen Yang, and Jingjing Zhang. When algorithms err: Differential impact of early vs. late errors on users' reliance on algorithms. *ACM Trans. Comput.-Hum. Interact.*, 30(1), March 2023.

[14] Max Langenkamp and Daniel N. Yue. How open source machine learning software shapes ai. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '22, page 385–395, New York, NY, USA, 2022. Association for Computing Machinery.

[15] Leona Lassak, Elleen Pan, Blase Ur, and Maximilian Golla. Why aren't we using passkeys? obstacles companies face deploying FIDO2 passwordless authentication. In *33rd USENIX Security Symposium (USENIX Security 24)*, pages 7231–7248, Philadelphia, PA, August 2024. USENIX Association.

[16] Patrick Yung Kang Lee, Ning F. Ma, Ig-Jae Kim, and Dongwook Yoon. Speculating on risks of AI clones to selfhood and relationships: Doppelgänger-phobia, identity fragmentation, and living memories. 7(CSCW1), Apr 2023. https://doi.org/10.1145/3579524.

[17] Zhuoran Lu, Dakuo Wang, and Ming Yin. Does more advice help? the effects of second opinions in ai-assisted decision making. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW1), April 2024.

[18] Takuya Maeda and Anabel Quan-Haase. When human-ai interactions become parasocial: Agency and anthropomorphism in affective design. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, page 1068–1077, New York, NY, USA, 2024. Association for Computing Machinery.

[19] Jaron Mink, Licheng Luo, Natã M. Barbosa, Olivia Figueira, Yang Wang, and Gang Wang. DeepPhish: Understanding user trust towards artificially generated profiles in online social networks. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 1669–1686, Boston, MA, August 2022. USENIX Association.

[20] Jennifer Mitzen. Ontological security in world politics: State identity and the security dilemma. *European Journal of International Relations*, 12(3):341–370, 04 2006.

[21] Frank Nagle and Daniel Yue. The latent role of open models in the ai economy. *Available at SSRN 5767103*, 2025.

[22] Helen Nissenbaum. A contextual approach to privacy online. *Daedalus*, 140(4):32–48, 10 2011.

[23] Duncan Pritchard. Why technology doesn't normally make you dumber, but agentic ai will. *International Journal of Human–Computer Interaction*, pages 1–11, 02 2026.

[24] J.H. Saltzer and M.D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.

[25] Téo Sanchez, Oleksandra Vereschak, and Ophelia Deroy. Mental models in human-ai interaction: Systematic review of empirical methodologies and guidelines. In *Proceedings of the 31st International Conference on Intelligent User Interfaces*, IUI '26, page 663–682, New York, NY, USA, 2026. Association for Computing Machinery.

[26] Beau G. Schelble, Christopher Flathmann, Nathan J. McNeese, Guo Freeman, and Rohit Mallick. Let's think together! assessing shared mental models, performance, and trust in human-agent teams. *Proc. ACM Hum.-Comput. Interact.*, 6(GROUP), January 2022.

[27] Katie Seaborn, Norihisa P. Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. Voice in human–agent interaction: A survey. *ACM Comput. Surv.*, 54(4), May 2021.

[28] Filipo Sharevski, Jennifer Vander Loop, Bill Evans, and Alexander Ponticello. (blind) users really do heed aural telephone scam warnings. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 74–92, 2025.

[29] Filipo Sharevski, Jennifer Vander Loop, Bill Evans, and Alexander Ponticello. "You creep! It really worked!": An empirical study of telephone scams with cloned familiar voices and trusted caller IDs. In *Proceedings of the 2025 New Security Paradigms Workshop*, NSPW '25, page 92–107, New York, NY, USA, 2026. Association for Computing Machinery. https://doi.org/10.1145/3774761.3774918.

[30] Mariarosaria Taddeo. Is cybersecurity a public good? *Minds and Machines*, 29(3):349–354, 2019.

[31] Christian Tarsney. Deception and manipulation in generative ai. *Philosophical Studies*, 182(7):1865–1887, 2025.

[32] Ken Thompson. Reflections on trusting trust. *Commun. ACM*, 27(8):761–763, August 1984.

[33] Jan-Philip van Acken, Floris Jansen, Slinger Jansen, and Katsiaryna Labunets. Who is the IT department anyway: An evaluative case study of shadow IT mindsets among corporate employees. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pages 527–545, Philadelphia, PA, August 2024. USENIX Association.

[34] Ge Wang and Roy Pea. Algorithmic autonomy in data-driven ai. *Commun. ACM*, 69(3):74–81, February 2026.

[35] Kevin Warren, Tyler Tucker, Anna Crowder, Daniel Olszewski, Allison Lu, Caroline Fedele, Magdalena Pasternak, Seth Layton, Kevin Butler, Carrie Gates, and Patrick Traynor. "better be computer or i'm dumb": A large-scale evaluation of humans as audio deepfake detectors. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, CCS '24, page 2696–2710, New York, NY, USA, 2024. Association for Computing Machinery.

[36] Rick Wash and Emilee Rader. Influencing mental models of security: a research agenda. In *Proceedings of the 2011 New Security Paradigms Workshop*, NSPW '11, page 57–66, New York, NY, USA, 2011. Association for Computing Machinery.

[37] Dominik Wermke. Understanding trust and security processes in the open source software ecosystem. In *Enigma 2023*, Santa Clara, CA, January 2023. USENIX Association.

[38] Alma 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., August 1999. USENIX Association.

[39] Rui Zhang, Nathan J. McNeese, Guo Freeman, and Geoff Musick. "an ideal human": Expectations of ai teammates in human-ai teaming. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW3), January 2021.

[40] Mary Ellen Zurko and Richard T Simon. User-centered security. In *Proceedings of the 1996 workshop on New security paradigms*, pages 27–33, 1996.