Navigating Autonomy: Unveiling Security Experts' Perspectives on Augmented Intelligence in Cybersecurity: Codebook

Other Research Data

Author(s):
Roch, Neele

Publication date:
2024-08-11

Permanent link:
https://doi.org/10.3929/ethz-b-000674517

Navigating Autonomy: Unveiling Security Experts' Perspectives on Augmented Intelligence in Cybersecurity
Complete Codebook

Neele Roch
ETH Zurich

Hannah Sievers
ETH Zurich

Lorin Schöni
ETH Zurich

Verena Zimmermann
ETH Zurich

Abstract

The rapidly evolving cybersecurity threat landscape and shortage of skilled professionals amplify the need for technical support. AI tools offer great opportunities to support security experts by augmenting their intelligence and allowing them to focus on their unique human skills and expertise. For the successful design of AI tools and expert-AI interfaces, however, it is essential to understand the specialised security-critical context and the experts' requirements. To this end, 27 in-depth interviews with security experts, mostly in high-level managerial roles, were conducted and analysed using a grounded theory approach. The interviews showed that experts assigned tasks to AI, humans, or the human-AI team according to the skills they attributed to them. However, deciding how autonomously an AI tool should be able to perform tasks is a challenge that requires experts to weigh up factors such as trust, type of task, benefits, and risks. The findings informed a decision framework that enhances understanding of the interplay between trust in AI, especially influenced by its transparency and different levels of automation and autonomy. As these decision factors affect the adoption of AI and the success of expert-AI collaboration in cybersecurity, it is important to further investigate them in the context of experts' AI-related decision-making processes.

1 Introduction

This is the full codebook, that was developed during analysing interviews with 27 cybersecurity experts. To analyse the interview data, we followed a grounded theory approach [3, 4], which is suitable when little is known about the topic and allows synthesizing qualitative interview data and generating research assumptions and frameworks [1]. The interview guideline aimed to elicit diverse responses on the participants' perceptions initially, but provided guidance to experts, prompting them to evaluate AI in the context of specific cybersecurity tasks. Therefore, responses focused on similar tasks and allowed the emergence of consistent patterns.

We used the central element of grounded theory, ongoing memoing, in the transcription and during all coding phases, to capture impressions and ideas. Memoing describes the process of recording thoughts, analytical insights, decisions, and ideas in relation to the research process [3]. During the coding phase, we added the technique of diagramming [3], i.e., creating visual representations of interrelations between codes, to support the development of the categories and their relationships and interactions. The coding process was structured as follows: the initial open coding phase aimed at initial codebook development, where the first five interviews were coded with a line-by-line approach. Once an understanding of underlying themes in the data was developed, line-by-line codes were transformed into incidents.

The intermediate coding process focused on axial coding. Strauss and Corbin defined axial coding as *"a set of procedures whereby data are put back together in new ways after open coding by making connections between [and within] categories"* [3] (p.96). Through diagramming and generating situational maps, the relationships of the arising themes and their contexts were captured [2]. During axial coding, two researchers went through the interviews and developed a situational map depicting the codes until no new codes or relationships could be added, and the situational maps were considered complete.

In the final coding phase, the codebook derived from the situational maps was transferred to the coding software MAXQDA (v24.1.0) [5]. The interviews were then coded topic by topic, and interview sections discussing the same topics were compared between participants to enrich themes and provide different variations of one topic and respective codes.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2024.
August 11–13, 2024, Philadelphia, PA, United States.

2 Codebook

Category	Description	Example
1. Human capabilities		
1.1 Strategy + Assessments	Humans are able to strategize, plan and assess as a unique human capability.	"[...] And, of course, strategic thoughts. So if I wanted to put myself in an attacker's shoes. What are the attacker's motivations, what do they want to achieve? Humans are certainly much better at generating such an understanding than an AI." (ME8).
1.2 Common sense/intuition	Humans have common sense and intuition, the unique capability to judge without a formal or logic reasoning.	"I call it common sense. Humans are still able to link information in a fairly unique way. I think human learning, human understanding, has not yet been overtaken by an AI." (ME9).
1.3 Creativity	Humans are creative.	"That is something I would say the human [is good at], this creativity." (ME3).
1.4 Communication: Empathy + Emotion	Humans have the ability to be empathetic, understand other people's needs and interpersonal relationships.	"And I think that is where humans are good at grasping feelings; essential, but perhaps not technically stored information." (ME12).
1.5 Moral compass	Morality is the belief that some behaviour is right and acceptable and that other behaviours are wrong.	"Either we are perhaps not yet ready as a society or as human beings, or perhaps a technology is not yet ready to really make decisions. There are also ethical and moral aspects. For me, the best example is always autonomous driving. [...]" (ME13).
1.6 Understanding	The human is able to understand the bigger picture.	"Recognizing connections where, at first glance, may be no connections at all." (ME7).
1.6.1 Context	Humans are able to justify or understand the context for tasks and decisions and integrate non-task related knowledge and factors into the decision.	"I mean, I think today humans can still bring in context in a way that AI cannot. [...]" (ME6).
1.6.2 Goal	Humans are able to assess situations in regard to a specific goal or desired outcome.	"That the human really takes over and then finishes analysing the incident [...]." (ME9).
2. AI Capabilities		
2.1 Limitations AI	The expert's considerations of the AIs limitations.	"[...] However, I see some limits here, especially with automated response. [...]" (ME22).
2.2 Analysing	AI is described to be good for handling and analysing big data, monitoring, filtering, detecting anomalies, and verifying and validating formal criteria continuously.	"Behavioural-typical analysis or behaviour-based analysis of how something behaves, how something moves. Is it usual? Is it unusual? I think that's actually exactly where AI can already be used very well today." (ME22).
2.2.1 Big data	AI is able to draw on huge amounts of known data.	"Things that a person may not be able to grasp or process due to the sheer volume. Where the human might not be able to bring sufficient focus and resolution to the topic, the topic of big data, looking for the needle in the haystack, for example." (ME9).
2.2.1.1 As knowledge base	AI is able to draw on vast amounts of data.	"And it would be very important for me that it [AI] can access the data directly. I know that it is already possible in some cases to say, Okay, in this or that process it sees this or that challenge and that it then tells me the process straight away." (CE24).
2.2.1.2 Handling	AI is able to process large amounts of data.	"Wherever routine tasks have to be carried out and, most importantly, large amounts of data have to be processed." (ME12).
2.2.2 Detecting	AI is able to detect anomalies, from a baseline normal pattern.	"I think everything that involves data analysis. Pattern recognition, but also the detection of anomalies. Based on a database." (ME8).

Continued on next page

Table 1 – *Continued from previous page*

Category	**Description**	**Example**
2.2.3 Continuity	AI can continuously execute its work without interruptions.	"I think AI is generally good for things you want to reproduce uniformly. It has a certain continuity in ensuring the processes." (ME5).
2.2.4 Filtering	AI can filter out unimportant noise.	"So, I think filtering out and checking false positives will most likely be a highly automated task." (ME13).
2.2.5 Monitoring	AI is good at monitoring data in real-time.	"And a second field of application I would see fit is to use AI in monitoring. In SIEM, or something like that." (ME2).
2.2.6 Requires formal criteria	As AI is based on algorithms, it needs a formal system to be able to work.	"I mean, AI is always as good as a) the algorithms, and b) the training material." (ME7).
2.2.7 Verifying and validating	AI can efficiently verify and validate information against a given or desired condition.	"In addition to writing code, you could also tackle configuration management, check firewall rules, check technical elements of configurations, etc. I think there would be a lot of added value there." (ME3).
2.3 Decision-support	AI is described to be good for preparatory tasks for the human decision-maker. This specifically means extracting, condensing, summarizing information, and mapping e.g., requirements with the status-quo.	"And the other is the actual incident, processing, analysis, etc., where [AI] can be very supportive. That brings us back to the point of making decisions, doesn't it? Working out the basis for decisions, yes, making decisions, no, maybe." (ME13).
2.3.1 Condensing	AI can make information accessible to humans by condensing it.	"That helps me a lot as a CISO. For example, I can say, Hey, briefly explain to me again, what does zero trust look like, how does passwordless work? What are the biggest current cyber threats? Of course, it helps me a lot in that respect." (ME26).
2.3.2 Extracting	AI can extract core information from bigger amounts of data or information.	"For example, if I give ChatGPT a task to find something for me, then it more or less has the whole internet as a source of knowledge at its disposal. I think AI is very good at collecting information in a compact form, at filtering it. To pick out the right information from a wide range of information available on the internet and provide me with what I have asked for." (ME2).
2.3.3 Providing Information	The AI can provide (targeted) information.	"For me personally, it is a decision support, a decision aid, and it can explain trade-offs." (ME13).
2.3.4 Summarizing	The AI can summarize large amounts of text.	"That means summarizing things, listing points that I can then reuse later." (ME7).
2.4 Generating + Enhancing	AI is described to be good for quickly generating knowledge, textual documents or a different perspective.	"A bit of this whole legal tech topic. Can't we automate the evaluation or drafting of certain documents? Well, I think [AI] can do that reasonably well when it's not hallucinating." (ME3).
2.4.1 Knowledge	The AI can be used to quickly access targeted information and provide knowledge to the human.	"[...] AI is not just theory, it is also practice, but a lot of theory. That means it might help me to revise an opinion, or it helps me to get additional material." (ME7).
2.4.2 Perspective	The AI can generate different perspectives through text.	"I can ask why it [AI] does certain things and that gives me a different perspective. [...]" (CE24).
2.4.3 Text	AI is used to generate texts of all kinds, e.g. documents, communication, etc.	"I am currently using AI very intensively to write policies, for example. [...]" (ME7).
2.5 Responding	AI is described to be good for responding to security incidents, or to questions.	

Continued on next page

Table 1 – *Continued from previous page*

Category	Description	Example
2.5.1 To incidents	AI is described to be good for responding to security incidents.	"[...] Or, let me say, when it comes to isolating clients, for example. Do we want an AI to do that on its own? Or does someone still have to check the case and say, Yes, it is okay. Regarding incident handling very much is already possible." (ME13).
2.5.2 To questions	AI is described to be good for responding to questions.	"[...] LLMs can take a question and provide a good outline, a very good memo, a very good resume. [...]" (ME6).
2.6 Translating	AI is described to be good for translating between different languages, but also different levels of expertise.	"[...] AI also enables me to translate [...] at a very high level, I would say." (ME7).
3. Tasks		
3.1 Technical support	The technical support that experts desired.	"In our security operations centre, where we also do detection and response, we often have to evaluate cases that are being reported, we have to go into the logs, we have to see what the procedure was, which kind of potential counterpart is involved? What is the intention behind it? What can be done about it? [...] These are things, for example, where we already have certain AI components in the solutions we use, but I think we can perhaps do more in the future." (ME22).
3.1.1 Routine tasks	Support with repetitive tasks that always follow the same or a similar structure.	"It does not just have to be AI, but where we hope or wish for support is of course in all routine tasks." (ME12).
3.1.1.1 Administrative tasks	Experts desire technical support for any administrative tasks.	" [...] Yes, and of course I would like to reduce all of this administrative, bureaucratic stuff. Gladly on technology if that is possible." (ME14).
3.1.2 Identify user from incident	Experts desired technical support to identify which user is affected by a security incident.	" [...] It is just that finding people requires a lot of technical effort [...]. There are lots of people with hacked devices [...]. Because we have an accumulation of tourists and so on, who then use our network and are compromised. And the GUI interface on which I have to search with lots of clicks and everything, [...] this is very, very tedious for me." (OE11).
3.1.3 Q&A chatbot	Experts would like to have support from a technical solution for standard questions.	"I think where we are actually getting to is the whole issue of structured analysis at the moment. Here are 20 legal texts. I will ask you questions about them." (ME3).
3.1.4 Complex constructs	Support with complex and/or time-consuming matters.	"[...] Yes, the ISMS is still somewhat static. We will probably also use a tool. And what we are currently thinking about, working on and carrying out our mini PoC, is the graphical representation of the dependencies [...]. Everything from the decrees, directives, and work instructions to security policies, and security design patterns at architecture level, right down to the solutions." (ME13).
3.1.4.1 Visualize	Experts desired support for visualizing complex constructs.	"Certainly, what would help us in the workshops would be if we had more visualizations." (CE24).
3.1.4.2 Structure and organize	Experts desired support for structuring, organizing and improving complex constructs.	"For example, we thought about whether we could use AI and metadata to structure this much better and get a better overview of our information." (ME21).
3.1.5 Knowledge and information gathering	Use of AI to gather information on topics efficiently, or to inform decisions.	"I mean, certainly, maybe one method for the whole thing, this whole information gathering, would be to use AI as a support." (CE24).

Continued on next page

Table 1 – *Continued from previous page*

Category	Description	Example
3.1.5.1 About (third-party) incidents and vulnerabilities	Use of AI to gather information on third-party vulnerabilities.	"Our organization probably has somewhat of a [...] range of third-party companies that support us, and they are not on the same security level as we are. And that means there are constant incidents. [...]. And I think it would be helpful for the industry if there were methods to automate and standardize the due diligence of such third-party companies, or even to scan or analyse such a company." (ME17).
3.1.6 Defensive cybersecurity	Use of AI to improve defensive cybersecurity capacities.	"Now, of course, the big focus in the security operations centre or cyber defence centre is supporting the management of false positives, alert fatigueness. This is a classic area where many vendors are now jumping on the bandwagon and trying to sell AI." (ME7).
3.1.6.1 Automatic response	Use of AI to automatically respond to detected anomalies.	"But there are other things, such as system monitoring. For example, if I see this or that error occurring, the automatic intervention and [...] running a script to correct it. And then only involve the human if the script somehow cannot solve the problem." (ME25).
3.1.6.2 Continuous (pen-)testing	Use of AI to automatically penetration-test.	"[...] You often only have a snapshot when it comes to cybersecurity. Especially when it comes to pen-testing. But also there, it would be nice to have a mechanism which results in similar quality or manual testers that are brought in more regularly." (ME23).
3.1.6.3 Filter and prioritize alerts and incidents	Use of AI to filter and prioritize security related alerts.	"[...] You get countless reports on how well or badly positioned you are. And when you look at the masses, you do not even know where to start. The question is how you can fish out the five or ten percent that you need to take care of." (ME23).
3.1.6.4 Threat intelligence	Getting insights on current threats and descriptions of how this might affect the respective organisation, or on the urgency of a threat or vulnerability.	"And the other thing, of course, is the threat landscape, where you have to process all the threat intelligence information that you receive internally. Which today is a lot of manual work, at least for us. There is certainly a potential there, I think." (ME20).
3.1.6.5 Automatically analyse incoming mails	Use of AI to automatically analyse mails for phishing.	"[...] And for me, it is actually more important that [...], either when a user clicks on or when a user reports a phishing E-mail, we actually have a relatively high level of automation there, which goes so far as that the system actually takes the E-mail, analyses it, checks whether other employees have received the same E-mail, and then, if necessary, flags it or puts it in quarantine or, perhaps better, deletes it." (ME16).
3.1.6.6 Monitoring and detection (profiling)	A technical solution could be used to streamline the monitoring of users and entities and analyse available data in real-time and around the clock.	"And a second field that I can picture the use of AI in incredibly well is monitoring. For example, in the area of SIEM or something like that. Detecting abnormal attacks, abnormal behaviour, when networks might have anomalies, etc. I think that is what AI is incredibly good at." (ME2).
3.1.7 Develop policies, frameworks, and guidelines	Technical support could be used during the life-cycle of policies, to set up, improve, write, and review policies and similar documents.	"For one thing, I think you could use text-based AI for dry work as defining some policies and processes." (ME2).

Continued on next page

Table 1 – *Continued from previous page*

Category	**Description**	**Example**
3.1.8 Analysing reports	Technical support could be used to analyse reports or do survey analyses.	"[...] We have a lot of phishing reports coming in and then that is something that is a tweaking, and having a quality filtering is also something that is hard to do. [...] With AI, we are making well done phishing and, you know, are trying to assess whether it is phishing. So this is something that I think also is a big opportunity." (ME1).
3.1.9 Communication facilitator	Technical support could be used to streamline the communication between different parties, and to communicate technical information at various degrees of technical expertise, e.g. non-tech-savvy people versus technically knowledgeable people.	"And also bridging the gap between different technical levels because we have a lot of people who come into the necessity of compliance or implementation of security controls, but they are not technical. You need to meet people on different levels. And that is something that is not done well by humans who have limited time and patience." (ME1).
3.1.9.1 Generator	Use of AI to generate texts, images etc.	"Of course, visualizations of some of the slides are also a lot of work. Also making the slides super engaging for the customers." (CE24).
3.1.9.2 Translation	Use of AI to translate between different languages.	"What I would wish for, what would be helpful in an international environment, of course, is textual translation." (ME18).
3.2 Lack of resources	The lack of resources' experts mentioned.	"[...] Specialized resources would actually be needed for all the areas mentioned. [...]" (ME3).
3.2.1 Tools	There is a lack of specialized tools for one specific task or area.	"So what is currently holding us back [...] is that we still do not have an enterprise architecture tool. [...]" (ME9).
3.2.2 Awareness	There is a lack of awareness for cybersecurity in the organization, which leads to a lack of other resources.	"The challenge lies not so much in security itself, but in the day-to-day business. Where people have guidelines to implement. So in terms of security, I would say we have clear guidelines, but that people apply them or that they know who is responsible for what. That is a big challenge." (ME23).
3.2.3 Focus	There is a lack of focusing on what is important. This could also include having to find the balance between what is usable and what is secure.	"We are working on a lot of fronts, and that naturally slows down progress when you are working on so many projects at the same time." (ME9).
3.2.4 Time	There is a lack of time to complete all tasks.	"That is just a challenge from a time perspective, not a budget perspective. That they can teach themselves enough to keep up to date." (ME8).
3.2.5 Workforce	There is a general lack of workforce; there are not enough employees to complete all tasks. This does not specify needing specific specialized skills.	"The financial, human and technical resources that are actually lacking and therefore make the overall framework quite unstable." (ME3).
3.2.6 Skills	There is a lack of skills available to ensure the protection of an organization's assets. Additionally, this could mean that the existing human employees do not have the capacities to do further training and develop their existing skills to keep up to date with recent developments.	"What is certainly always a challenge when there is a lot going on is further training. Maintaining the assurance. That is a challenge. It goes beyond normal working hours to keep up with developments in the field of cybersecurity." (ME8).

Continued on next page

Table 1 – *Continued from previous page*

Category	**Description**	**Example**
3.2.7 Monetary	Resources are lacking in the monetary area, specifically the lack of budget. Clients or stakeholders only attribute a specific amount of budget that is not sufficient to have, e.g., enough manpower or acquire appropriate infrastructure to ensure information security.	"[...] and on the other hand, of course, it also costs more and more in terms of wages." (ME20).
4. Experts		
4.1 Hopes	The hopes experts have regarding the deployment of AI in cybersecurity, and in general.	"That it [AI] continues to develop and naturally brings maximum added value." (ME5).
4.2 Worries	The worries experts have regarding the deployment of AI in cybersecurity, and in general.	"In such situations I am relatively cautious. On the one hand, in terms of what I am being told, because the machine has perhaps analysed me and knows my strengths and weaknesses and is perhaps playing my weaknesses against me. [...] I say a healthy dose of respect is necessary." (ME13).
4.2.1 Misinformation	Experts worry about getting false information for an AI.	"The checking of information. I have a very critical view on it. How can we as users judge whether what is going on is plausible? Has it been changed? Has it been manipulated? I see this as a critical challenge in the overall context." (ME26).
4.2.2 AI developing consciousness	Experts worry about AI developing a consciousness and therefore might go from a tool to having its own agenda.	"The keyword being consciousness. If at some point a system were actually to develop something like consciousness, that would be something which results in you having to reassess a lot of things. [...] Then, of course, the risks would be completely different." (ME5).
4.3 Expert responsibilities	A summary of the reported responsibilities of the experts.	"[...] I am part of reviewing policies and then designing controls. [...]" (ME1).
4.4 Expert tasks	These are the tasks the experts talk about needing to complete to fulfil their responsibility.	"The tasks are, of course, different. One is of technical nature, like setting up and checking whether the right users have the right rights. Then, of course, checking accounts when people change their jobs or leave the company. [...]" (ME4).
4.5 Expert roles	The roles the experts take on in their profession.	"So my role is Chief Information Security Officer in the private sector CISO." (ME3).
5. Expert-AI interface		
5.1 Interface	Descriptions of what the interface between experts and AI should look like.	"[...] An AI should feel natural, so that you enjoy using it, and it does not feel strange." (ME10).
5.1.1 Privacy concerns	The experts address concerns related to safekeeping of data, and maintaining the organisation's or even their personal privacy	"How can we retain control? How can we prevent our intellectual property from simply being passed on for free, so to say? For example, We do not want all our articles to be freely available anywhere." (ME26).
5.1.2 Context-awareness	The AI needs to understand the context, the goals, and the organization.	"The AI must know the context of the company. [...] It has to be context-based. It has to fit into the company." (ME25).
5.1.3 Communication	The desire for or anticipated communication between an expert and an AI system. This can include descriptions of the mode or what is deemed desirable or undesirable.	"In terms of LLMS, I see a great added value with regard to user interaction." (ME9).

Continued on next page

Table 1 – *Continued from previous page*

Category	**Description**	**Example**
5.1.4 Availability	This code describes the availability that experts describe for the AI system and its use, e.g. the desire for a constant availability, the system to be invisible in their workflow and similar descriptions.	"Yes, I think the presence is important. You know, so that I almost always have it to hand. That [...] I can access it at any point in time. No matter whether that is in the morning, whether something comes to mind at night [...]. The availability of AI." (ME14).
5.1.4.1 Integration	How the integration of AI should look like.	"[...] I do not see a virtual butler here who takes things off my hands, I see it being embedded in different processes, don't you?" (ME17).
5.1.5 AI-to-expert	This code describes aspects of the interface that are directional, specifically AI to expert, e.g. the AI alerting the expert proactively.	"[...] It depends on the context. It [the interaction] can be an alert, it can be a risk. Of course, it can also be a recommendation, or how you should proceed. Just a prioritization of what is possible with the technical data." (ME23).
5.1.5.1 Support	The AI should support the expert with his tasks.	"And then, once the whole thing is up and running, it goes in the other direction again. AI can then support the humans in their tasks and their recurring tasks, of course." (ME5).
5.1.5.2 Recommendation	The AI should recommend courses of action to the expert.	"[...] Instead, it [AI] should make a suggestion and always ask me whether it should really do it before making an independent decision. [...]" (ME2).
5.1.5.3 Report	The AI should report desired information to the expert.	"So communication would still take place [...] via the screen in the form of reports." (ME18).
5.1.5.4 Alert	The AI should alert the expert in the case that the expert's attention is required.	"The AI would have to have its own risk assessment mechanism, according to which it would then somehow delegate the tasks further to humans." (ME25).
5.1.6 Expert-to-AI	This code describes aspects of the interface that are directional, specifically expert to AI, e.g. experts being able to ask questions to the AI.	"[...] I actually think it [the interaction] should be as it is now, like what we see with AutoGPT and ChatGPT, or anything else, where you can just say, Hey, I just had a thought, give me some input on it, or I will throw in 20 legal texts and ask it about them. A selective cooperation, the integration in everyday tools." (ME3).
5.1.6.1 Ask questions	The expert wants to be able to ask the AI questions.	"I can ask questions and get enriched information, but preferably with an explanation. What, of course, would not be useful to me in our job is when I would simply get a statement. [...]" (ME12).
5.1.6.2 Request task execution	The expert wants to be able to request task execution by the AI.	"Of course, if an image analysis is able to analyse a problem and extract a question from it and perhaps even give me an answer in the same way, that would be very helpful." (ME9).
5.2 AI role	What the role of an AI is described as.	"As I mentioned before, I would like to have some sort of assistant or co-pilot, who can provide me with information, so to speak." (ME12).
5.2.1 Tool	An AI system is described as a tool.	"I used it [AI] for awareness trainings. [...]" (ME21).
5.2.2 Generator	An AI system is described to be used as a generator, the output could be text, images, documents, etc.	"Yes, so if I had to write some texts, for example, I would just say, Hey, please draft a text on this specific topic or prepare a presentation [...]". (ME22).
5.2.3 Assistant	An AI system is described as an assistant helping experts with specific tasks and assisting them in their daily workflow, etc.	"[...] You could think of an assistant. You could say, Hey, I am giving you these words or this specific search query, now help me find something." (ME17).

Continued on next page

Table 1 – *Continued from previous page*

Category	Description	Example
5.2.4 Co-pilot	An AI system is described as a co-pilot. A co-pilot helps the primary pilot operate, assists with various tasks and supports the primary pilot to accomplish their goal.	"I think, especially in the area of co-pilots, we already roughly know what that might look like. [...] Of course, we already see this in the area of software development, where it has been available for some time and I have a co-pilot to program a software. An important aspect there is that the software is programmed as securely as possible." (ME8).
5.3 Expert role	What the role of the expert is described as.	
5.3.1 Overseer	The expert takes the role of an overseer that checks, validates and approves or modifies the work produced by an AI to ensure a certain quality or standard.	"But in the end, I need to know about the topic, so that I can validate, whether what it [AI] is suggesting makes sense, or whether it is completely wrong. I mean, AI can also produce false correlations or misinterpret data. [...]" (ME22).
5.3.2 Decision-maker	The expert is the final decision-maker when AI and experts work together. The expert decides the best course and is at the same time also the one holding responsibility in those cases.	"I do not think that we are in a position to assign responsibility to an AI [...]. Ultimately, it is up to the humans to take responsibility." (ME27).
5.3.3 Provider/developer	The expert is described as the developer of an AI. This includes the curation of data for training, and the development of the model.	"I think, initially, it is probably the case that the human has assembled the AI, and has created it, so to say. So that the human also gives it certain abilities by configuring the algorithm accordingly." (ME5).
6. Autonomy		
6.1 Determination	The AI system is described as deterministic where one input always leads to the same output, or a predefined set of rules define the course of action.	"I am saying that, given the same parameters, a person would perhaps make a different decision if they were influenced differently emotionally at that moment. But since LLMs do not have emotions, I ask myself: Why do they come to a different conclusion?" (ME19).
6.2 Transparency	The AI system needs to be transparent and understandable, so that experts can understand how the AI came to a specific output based on the input that it got.	"I think that in order for me to fully trust an AI, it would have to be able to tell me why it came up with the result when I ask questions." (ME19).
6.3 Task type	Descriptions of what is important for the type of tasks affecting the autonomy.	"To combine specifications. You know, to formulate a directive, write it. But I would leave the implementation to the humans." (ME14).
6.3.1 Capability fit	Descriptions of the capabilities that fit the type of task and how this justifies the autonomy.	"[...] From this perspective, it would be very interesting with regard to the risk analysis to collect the information and then link it to the company's information and then to actually generate a complete picture with the focus points." (ME20).
6.3.2 Urgency	Descriptions of the criticality of the status quo and a certain assigned level of autonomy to the AI and how they influence each other.	"I mean, the criticality is not that high. You can make an autonomous decision, per se." (CE24).
6.4 Benefit	Description of whether the use of AI makes the work of an expert any less, or easier, or more efficient, and whether the expert and the respective company benefit from utilizing AI.	"AI will enable us to do things that we could not do in the past, simply because we can analyse data better, faster and differently." (ME22).

Continued on next page

Table 1 – *Continued from previous page*

Category	**Description**	**Example**
6.4.1 Accuracy	The accuracy plays an important role in the decision of autonomy. If the accuracy of an AI system is perfect, then the level of autonomy of an AI can be rather high; if the accuracy of an AI system is not perfect, which is the current ground truth, the decision regarding autonomy is based on different factors.	"Oh, it depends on the false positives and the false negatives. And on the impact of the false positives and the true positives. So I do not know about that." (OE11).
6.5 Trust	Trust plays an important role when experts decide how autonomous an AI system should be able to act. If their trust is low, they are not willing to let an AI system act autonomously; if the trust is high, they are willing to give away more control.	"If I create my own AI, so to say, then of course I would have greater trust than if it is a public or bigger one." (ME26).
6.5.1 Counter	Counts how many experts trust an AI, and how many do not.	
6.5.1.1 Distrust	Experts express not trusting an AI.	"Then I cannot trust that [AI]. No. You cannot." (OE11).
6.5.1.2 Conditional trust	Experts claim they trust an AI and at the same time express a condition or scenario in which they trust and in which they do not trust. Their trust is tied to conditions.	"I think with limited scope and having strong mechanisms of ensuring that the outcomes are correct, then there can be trust and reliance." (ME1).
6.5.1.3 Trust	Experts express trusting an AI.	"Yes, I can. Because we are very analytical, purely in terms of our job description. Of course, I generally trust systems, yes." (ME12).
6.5.2 Weaken	Factors that experts mentioned that would weaken their trust in AI.	"I believe that if I am then increasingly served with false information, take any AI translation tool as an example, if I were to realize that the content is not correct, that would weaken the whole thing." (ME27).
6.5.2.1 Misuse	Trust in an AI can be weakened by humans misusing it for the wrong purposes.	"And I mean, we all know what the big four or five providers want. They want to make money out of it. And how can we, as a society, retain control over a system that then has a massive impact on our daily activities, on our work, on our private lives?" (ME26).
6.5.2.1.1 Propaganda	Humans misuse AI for spreading false information.	"That would of course be an issue if it [AI] also spits out false topics, not in terms of logic, but really provides false information. Information that is actually fake, especially in this information environment, would of course fundamentally shake my trust and I would refrain from using it further." (ME26).
6.5.2.1.2 Social scoring and profiling	Humans misuse AI for social scoring or profiling, which results in a loss of trust.	"It goes a bit in the direction I already mentioned, excessive profiling of users. I'm really torn on that one. [...]" (ME9).
6.5.2.2 Negative experiences	Experiences of an AI giving results that do not make sense or are visibly wrong, hallucinations, and publicly reported data breaches or incidents can weaken the trust.	"I have just experienced too many things in my career that have simply gone wrong. Starting with virus scanners that picked up some wrong pattern, and then you come into work on Monday and all the rights are blocked." (ME21).

Continued on next page

Table 1 – *Continued from previous page*

Category	Description	Example
6.5.2.2.1 Security-related	Publicly reported data breaches or incidents. This includes phenomena such as adversarial prompts.	"I think losing trust is similar to seeing cybersecurity incidents at companies. [...]" (ME2).
6.5.2.2.2 Personal	Experiences of an AI giving results that do not make sense or a visibly wrong, or hallucinations.	"Well, when I ask ChatGPT to identify 50 AI-related topics in one article [...] and it does not make any sense at all. That happened to me recently. And it did not even give me 50 topics. [...] That certainly weakens my confidence." (CE24).
6.5.2.3 Black-box	Not understanding the model or its results can make the experts distrustful. Black-box models are not favourable to trust.	"It is a dilemma. On the one hand, of course, I cannot take every algorithm apart, I have to trust that it will produce valid and reliable results." (ME22).
6.5.3 Strengthen	Factors the experts stated would strengthen their trust in an AI.	
6.5.3.1 Regulation	National and international regulation can strengthen the trust in AI, as it holds developers accountable.	"I personally assume that this will also be regulated, somewhere at least in Europe, at some point. It will not last long. [...]" (ME26).
6.5.3.2 Provider	A trustworthy provide strengthens the trust in AI.	"The other is probably the principle of origin, depending on where it [AI] comes from, of course. Because there are, of course, certain manufacturers or sources that people think are more trustworthy, which makes you more inclined to trust the AI too." (ME5).
6.5.3.3 Transparent communication	Openly finding descriptions on how the model functions, understanding the infrastructure, and how data is processed can strengthen the trust in an AI.	"I think massive transparency. Where does the data come from? [...] What happens to my or our data? I think that is the really big sticking point. The transparency." (ME3).
6.5.3.4 Experience	Experiencing the AI give correct results can strengthen the trust in the predictions. Trust in AI is established based on the trial and error principle and observing the AI working correctly over time.	"When it turns out in retrospect that the decisions made were the right ones. That gives confidence. That gives security. [...]" (ME5).
6.5.4 Privacy	Considerations of how data is being processed inside the AI model plays a role in how and if an AI model is trusted. This might also apply to in-house AI models.	"That would also be the first thing that comes to mind when using AI. Is there compliance? Take the GDPR, for example. So in other words, to what extent is it legally compliant if I enter something into it [AI]? (ME2).
6.5.5 Transparency	Understanding the outcome, recommendations, and offered decisions by an AI model is an important factor for experts to trust the result. This includes how a model is trained, built, evaluated, but also how it has come to the results that it is presenting to the expert.	"Most of the time, when it comes to incidents, I would want a level four [of autonomy], so that I can keep track." (ME14).
6.5.5.1 Outcome	Experts describe the need to understand how an AI has come to some conclusion.	"It is often difficult to understand why the result of application A is the way it is, and the exact same case, only with nuanced differences, results in a completely different decision." (ME13).

Continued on next page

Table 1 – *Continued from previous page*

Category	**Description**	**Example**
6.5.5.2 Model	Experts want to understand the inner workings of an AI model. This includes how it is trained, what kind of model it is, or who the developers and providers are.	"So my specific limitation would simply be that it is difficult to compare the AI's that are implemented in the defence tools from different manufacturers, and that there is simply only a certain degree of transparency and visibility." (CE24).
6.5.6 Human oversight	For experts, trust seems to be tied to them being able to make the final decision. This might also include validating, verifying, modifying, or being able to overwrite an AI model's outcome or prediction. The experts trust an AI to the extent to which they can still intervene, and it cannot act completely autonomously.	"In both cases, it is up to me, who perhaps has more experience, to review it again before I pass it on to the customer. And it is precisely for these reasons that [...] from a legal and logical perspective, we have a four-eyes principle." (CE24).
6.5.6.1 Pre-set human defined rules	The expert explains that they want the AI to act in a specific predefined way for specific tasks every time.	"That I can say, so to speak, if this or that happens, then I will agree that the machine will respond in this or that way. [...]" (ME22).
6.5.6.2 Validation	The expert mentions the need to validate the correctness of the AIs outcome.	"But if we are taking it further now, and think about distribution. Then we are talking about the government [...]. Then I would say, No, I would like to look over it first." (ME3).
6.6 Risk	The considered consequences, implications, and their respective criticality that result from a certain assigned level of autonomy to the AI.	"So, if I listen to what I am saying, then it is actually always the case that the more risk involved in an action, the less I would trust an AI at the moment." (ME25).
6.6.1 Liability	The experts report liability issues that are associated with the outcomes of autonomous AI.	"[...] And the second reason is, of course, liability. This is a moral and ethical dilemma. I cannot hold a machine liable. [...]" (ME3).
6.6.2 Impact	How the possible impact of the executed reactions influences the choice of the level of AI autonomy. The impact may be minor or major. The severity of an impact might also be discussed. Different considerations for how justifiable the impact of consequences may be might also be discussed.	"I think it really comes down to the potential damage that could be caused by the decisions, and the scope of the decision. I think that is what it is. I would say when it comes to smaller matters that can easily be corrected, I would give a very high level of autonomy." (ME12).
6.6.3 Reversibility	How the reversibility of the executed reactions influences the choice of the level of AI autonomy.	"[...] Yes, one of the decision criteria for me is to what extent can I reverse the decision, if the human realizes that it was not correct for some reason?" (ME9).
6.7 Level unspecified	The use of AI with a certain degree of autonomy is mentioned, but no explicit level of autonomy is specified.	"Perhaps an interesting approach is if you take a self-declaration of risks, and you have to provide evidence on how certain things have been done, and how you came to your risk assessment, that an AI could be doing such a review of the evidence." (ME23).
6.8 Level 5	Level 5 of autonomy: The AI carries out actions independently and does not inform the expert.	"Employees from similar departments can probably be given similar access. Of course, this can be fully automated. Definitely." (ME4).
6.9 Level 4	Level 4 of autonomy: The AI carries out actions independently and informs the expert about the actions.	"Well, regarding certain evaluation of log files, you could really go to the extreme [level]. Nevertheless, I think I would probably start with the second most extreme, where the human still gets certain information." (ME4).

Continued on next page

Table 1 – *Continued from previous page*

Category	**Description**	**Example**
6.10 Level 3	Level 3 of autonomy: The AI acts unless the expert vetoes.	"Storage management, that could be something. [...] That could even take place at level three." (ME21).
6.11 Level 2	Level 2 of autonomy: The AI offers a decision and acts only after the expert approves.	"The AI sets an amount and proposes whether there is a bounty. And a human then checks before the payout." (ME14).
6.12 Level 1	Level 1 of autonomy: The AI suggests options and the expert decides.	"Regarding risk management, I would say as that is about taking responsibility, I would give it a level one." (ME14).

Table 1: Codebook

References

[1] Melanie Birks and Jane Mills. *Grounded theory: a practical guide*. SAGE, Los Angeles London New Delhi Singapore Washington DC Melbourne, 3rd edition edition, 2023.

[2] Adele E Clarke. Grounded theory: Critiques, debates, and situational analysis. *The SAGE Handbook of Social Science Methodology*, pages 423–442, 2007.

[3] Juliet M. Corbin and Anselm Strauss. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1):3–21, March 1990.

[4] Anselm L. Strauss and Juliet M. Corbin. *Basics of qualitative research: techniques and procedures for developing grounded theory*. Sage Publications, Thousand Oaks, 2. edition, 2003.

[5] VERBI Software. MAXQDA, 2024. https://www.maxqda.com/.