Social Transparency in Network Monitoring and Security Systems

Daria Soroko
darsorok@uni-bremen.de
University of Hamburg
Hamburg, Germany
University of Bremen
Bremen, Germany

Gian-Luca Savino
gian-luca.savino@unisg.ch
University of St. Gallen
St. Gallen, Switzerland

Nicholas Gray
nicholas.gray@uni-wuerzburg.de
Julius Maximilian University of Würzburg
Würzburg, Germany

Johannes Schöning
johannes.schoening@unisg.ch
University of St. Gallen
St. Gallen, Switzerland

ABSTRACT

System administrators (sysadmins) are key to keeping computer networks safe. As networks grow in size and complexity, partial workflow automation with the help of AI has been introduced to assist them. However, AI-aided tools often lack transparency, which may lead to the sysadmin's reluctance to use the new software, slower response time in case of a security breach, and increasing errors. Related work suggests that the concept of social transparency (ST), when applied to the IT-security context, enables peer support and could provide the missing knowledge to the user facilitating explainability of the system and improving human-AI trust. In this paper, we investigate the profile of sysadmins and confirm that ST can indeed yield benefits for them but only when coupled with relevant contextual information and only when it adheres to the sysadmins' quality standards. Finally, we contribute design recommendations for incorporating ST into the existing workflows of sysadmins.

CCS CONCEPTS

• **Networks** → **Network security**; • **Human-centered computing** → **Collaborative and social computing theory, concepts and paradigms**; Scenario-based design; User centered design; **HCI theory, concepts and models**; Empirical studies in HCI; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

network monitoring, artificial intelligence, automation, explainable AI, social transparency, system administrators, trust

ACM Reference Format:
Daria Soroko, Gian-Luca Savino, Nicholas Gray, and Johannes Schöning. 2023. Social Transparency in Network Monitoring and Security Systems. In *International Conference on Mobile and Ubiquitous Multimedia (MUM '23), December 03–06, 2023, Vienna, Austria.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3626705.3627773

MUM '23, December 03–06, 2023, Vienna, Austria

ACM ISBN 979-8-4007-0921-0/23/12.
https://doi.org/10.1145/3626705.3627773

1 INTRODUCTION

With an increasing number of services moving online, computer networks are growing in size and complexity [13, 23, 75, 83]. As a result, operating, monitoring and securing such large networks becomes a challenge for system administrators (sysadmins) and network operators (treated here as a subcategory of sysadmins). At the same time, the number and severity of attacks have increased dramatically in recent years [34, 48, 85]. A central and often difficult task for sysadmins is to have a full understanding of the network. For this, the required information is commonly gathered and visualised by a Network Management System (NMS) [44]. Yet, usability issues in existing software solutions and growing demands for broader knowledge and skills in the domain put a lot of pressure even on seasoned professionals [24]. In addition, the passing of knowledge and experience from competent experts to novices is not commonplace and depends on the enterprise culture of an individual company. Naturally, there comes the need for a system solution that could address these user needs: 1) Assist in securing large networks efficiently, 2) provide greater usability and user experience to sysadmins, 3) support knowledge acquisition by the admins to resolve issues in the network (both from external resources and from experienced peers).

Intelligent AI-based user interfaces promise to fulfil some of the stated needs by automating parts of the sysadmins' workflow, yet in reality often fall short in accomplishing the goal. The adoption of such systems is well underway [16, 22, 53, 75, 84], with more companies choosing to invest into advanced AI-based software solutions [27], such as next-generation firewalls (NGFW), network detection and response software (NDR), security information and event management systems (SIEM) etc. [1–3, 26, 39]. Such products focus on automatic threat detection through the analysis of devices, users and their traffic, and prioritise security alerts to allow sysadmins to efficiently resolve conflicts and breaches within the system. At the same time, transparency of these tools' decision-making processes is often sacrificed for the sake of convenience and efficiency, compromising the overall user experience. Based primarily on opaque AI algorithms, such as neural networks [22, 30, 81], these systems make it difficult to understand how they achieved their results [9, 42, 63, 76, 89]. This may lead to the user's inability to form a clear mental model of the system and decrease trust in its output [49, 55, 70, 77, 100]. Trust is an important factor, especially

in high-stakes scenarios: In the IT-security context, if the system is compromised, the sysadmin in charge of the network is likely to bear the brunt of responsibility for any losses associated with the breach. As a result, the users might fall into one of the two extremes: Either distrust AI completely and refuse to adopt the system entirely; Or choose to over-rely on the system, especially when overwhelmed by the number of tasks at hand [60]. Both extremes are especially undesirable in the context of network security. The lack of understanding of the system contributes to poor usability. The lack of transparency about the system and its output hinders knowledge acquisition by sysadmins, and consequently its distribution. Therefore in this paper, we investigate how the usability and transparency can be improved to enable knowledge acquisition by sysadmins.

A recent contribution in the field of Explainable AI (XAI), the social transparency framework (ST), could help to alleviate the aforementioned concerns. Its authors, Ehsan et al. [32], propose a more holistic approach to explainability by giving the users knowledge about their peers' past interactions with AI system. They argue that the framework has the potential to increase transparency and trust in the system, and provide other benefits to the users, also in the IT-security domain. Motivated by the approach of Ehsan et al. [32], we investigate:

RQ: If and how the implementation of social transparency can improve usability and transparency in AI-based systems to facilitate knowledge acquisition for system administrators?

To answer this research question, we conducted an expert study involving 12 system administrators. They were presented with a speculative design scenario involving a mock-up of an AI-based intelligent firewall. Following the interaction with the scenario, participants' responses were collected, thematically coded, and subsequently analyzed to derive meaningful insights. We found that system administrators, or sysadmins, are generally skeptical towards AI. This skepticism underscores the fact that trust in AI-based systems must be earned and is not automatically granted. Furthermore, sysadmins expressed a clear need to understand how these AI-based systems operate. Importantly, we also discovered that sysadmins can significantly benefit from social transparency, enhancing their interaction with AI-based systems.

Overall, in this paper, we test the ST framework in the IT-security context. We confirm the results of the related work and further extend them by contributing (1) an investigation of the profile of our target audience and their attitudes towards AI-based systems with a group of IT-security experts. (2) We investigate the sources that sysadmins currently use to gather contextual information when solving new problems. (3) We test the applicability of the ST framework to the IT-security domain in a dedicated study. Finally, (4) we propose design recommendations for incorporating ST into existing workflows of system administrators to facilitate explainability of AI-based systems and ameliorate some of the issues associated with this type of automation. The recommendations are based on the knowledge about how sysadmins solve the problems currently and based on the feedback they provided on the framework during this study.

2 RELATED WORK

Communication networks often lie at the core of several enterprise workflows and offered services, hence constituting a vital part of the corporate infrastructure. To ensure a smooth operation, these networks are configured, maintained and secured by system administrators (sysadmins), who are skilled and in some cases, highly specialised IT personnel. However, the tool ecosystems (f.ex. NMS, SIEM) they commonly use, are complex and prone to human error [29, 57, 99]. As shown by Maxion and Reeder [66], intelligent user interfaces and user-centric design can help to reduce the likelihood of errors, even for complex security tasks [29, 57, 99], and to aid sysadmins in handling large volumes of network data. To that end in recent years, advanced ML marketed as AI, has been incorporated more frequently into all aspects of communication networks, primarily to cope with the ever increasing amounts of processed data and its complexity [13]. New concepts like self-driving networks [35] are also under development and promise to provide their users a real-time control over the network to ease the management efforts by operating in closed loops without human intervention. And while AI-driven software solutions are becoming more commonly incorporated into the sysadmins' daily workflows, their usability, effectiveness and acceptance by the users have so far received little attention from the Human-Computer Interaction (HCI) community.

2.1 HCI in IT-Security

HCI topics, such as usability, have been mostly overlooked in the IT-security domain in general [24, 82, 87, 90]. Some early research published between 2001 and 2007 investigated sysadmins' user profiles, their tools and practices [10, 12, 43, 47]. However, when it comes to the more recent research, we encountered very few studies pertaining to the subject. Sysadmins as a user group remain largely neglected in HCI [90]. Several studies examined the UI design, use and effectiveness of command line interfaces (CLIs) and graphical user interfaces (GUIs) [94] and their conceptional integration to improve performance [71, 72]. User-centred approaches to designing and building visualisations for IT-security data were explored by McKenna et al. [69]. A different study investigated sysadmins' existing workflows and strategies they use to resolve conflicts in the network [98]. The admins' behaviour and attitudes towards updates of corporate software was the subject of a more recent paper by Tiefenau et al. [87], while Dietrich et al. focused on the users' perspectives regarding security misconfigurations [29]. Another study conducted a literature review on usability of firewall configurations [93], however discovered the "lack (or even absence) of usability evaluation or user studies to validate the proposed models" [93]. The vast majority of papers we have encountered, however, dealt primarily with novel methods for visualising the network activity and status [7, 31, 67, 68] including vulnerability analysis [5] and firewall rules configuration [54], as well as new ways of detecting and displaying anomalous and malicious behaviour in the network [21, 40, 41, 61, 62, 102]. Overall, not only is there limited research pertaining to usability of network monitoring and security systems, the few publications that offer usability guidelines and UI design principles [24, 82, 90, 96] do not touch on the subject of requirements for AI-based systems.

The effects of the existing HCI research gap in the domain are further aggravated by the new challenges that the incorporation of AI introduces to the field. While automating a lot of tasks, these

systems often merely display their findings to a sysadmin, who is burdened with the responsibility of the final call to action [56] without understanding of the underlying system mechanisms and information about them to support the user's decision. As a result, in case of failure, the system lacks explainability which leads to users having little idea about the source of the issue and ways to fix it, and thus, are less likely to trust the system's output in the future [58] or to rely on it.

2.2 Explainable AI (XAI)

An emerging discipline of explainable AI (XAI) aims to address some of the transparency and trust concerns across different domains. We believe, methods derived in XAI can be applied to IT security as well. XAI is a fairly new interdisciplinary field that aims to bring the background algorithmic processes of AI-based systems to the forefront. The goal is to help to empower the users of such systems with knowledge and understanding about the system and its limitations, thereby establishing trust in human-AI assemblages and providing control over the system and its results. No agreed-upon definition of XAI exists as of yet [33].

XAI puts significant emphasis on the ability of the system to explain its decision-making processes to the user. The majority of studies so far have been proposing post-hoc explanation strategies to help glean insights from the black-box models that are commonly used in AI-based software [32, 50, 64, 73, 86]. However, an increasing number of researchers are making an argument in favor of a more holistic approach to explanations that incorporates all stages of the decision-making process including the collection and labelling of the data used to train the AI model [6, 32, 33]. In addition, XAI covers a variety of factors that contribute to the users understanding of an AI system and their trust in it, such as types of explanations appropriate for different user profiles, types of human-AI interaction and levels of respective autonomy in human-AI teams, the concept of trust, its various manifestations and factors influencing its development [18, 49, 60, 101], and much more. In the process of reviewing existing literature, we also discovered three publications concerning XAI specifically in IT-security context and focusing on explaining the system's results in intrusion detection scenarios [65, 86, 92].

And while XAI helps to investigate the more algorithm-centred side of the human-AI interaction, it falls short of including contextual information such as the socio-organisational context that is usually not included in AI models [32]. Social context in the form of peer support is an important resource for system administrators even outside of the AI-mediated scenarios as it aids the admins in decision-making, when their own knowledge and experience are not sufficient [24]. However, AI automation in its current form adds a new layer of uncertainty and opacity. It makes the availability of social context imperative to the user's ability to make decisions in novel situations, to provide opportunities for sharing knowledge among peers and a sense of shared responsibility.

2.3 Social Transparency Framework

There are different approaches to integrating social context in digital systems. Social navigation is one of the most established topics in social computing that takes the natural human tendency to follow other people's cues when feeling lost [24, 28]. One could think of it as a form of collective intelligence. The concept was originally inspired by human behaviour in the physical world and applied to information spaces on the web. However, the term has evolved to not exclusively refer to the act of actual navigation but to also signify a type of decision-support mechanism [28]. Social navigation is the way the system administrators usually engage with social context in daily work. In practice, for sysadmins this often means consulting colleagues and online forums, such as Stack Overflow, to find remedies for the technical issues when facing them for the first time.

ST by Ehsan et al. [32], as we see it, is the latest iteration of social navigation. What sets it apart is its focus on AI-aided decision-making. ST is a framework that concentrates on using the external contextual knowledge about one's peers' interactions with AI to increase the explainability of AI-based system and to aid the user in making their own decisions. The approach is meant to address, among other things, the issues of trust and lack of understanding in the AI-supported decision-making processes, provide means to incorporate external human knowledge otherwise not included in the AI model and share it among the peers. The approach could yield great benefits for system administrators in particular. In the context of network monitoring and security, ST could help alleviate the issue of missing knowledge by consolidating information about the users' past experiences with the system in one location and disseminating the collective knowledge among peers. In doing so, the users could have a form of legacy system to support their decision-making in novel situations.

More importantly, ST could help the users to form a clear model of the system's processes and to cultivate what Lee and See called "appropriate trust and reliance" on automation [60]. In their seminal work, the authors argue that trust and reliance have a dynamic closed-loop relationship where trust influences reliance and reliance affects trust in an automated system. Lee and See also argued that instead of designing for greater trust, it is more sensible to design for appropriate trust (and reliance) where the user can make a more objective assessment of the system's capabilities and decide when and to what extent they should trust/rely on the automated system in the process they called "calibration" [60]. To help the users to make such an assessment, the authors stressed the importance of clearly conveying the capabilities of the automation to the user by bringing its operation to light. However, as discussed earlier this can be challenging due to opacity of AI algorithms. At the same time, switching to a more transparent algorithm may be undesirable in the IT-security domain as it may result in a decreased performance/effectiveness and security of the system. This is where ST could prove beneficial by providing information about the peers' past interactions with the system and, therefore, increasing the explainability of the system. This, in turn, could help the user to form a clear mental image of the system's capabilities and cultivate appropriate trust and reliance on it.

According to Lee and See [60] in addition to trust, self-confidence also influences the users' reliance on automation. Self-confidence is the person's assessment of their own capabilities to perform a task. The authors state that: "When operators' self-confidence is high and trust in the system is low, they are more inclined to rely on manual

control. The opposite is also true: Low self-confidence is related to a greater inclination to rely on the automatic controller" [59, 60]. It is worth noting that people on average have a tendency to be overconfident in their abilities [37, 38, 59], and access to more information may increase self-confidence even further [74] as cited in [59]. And when the user's self-confidence doesn't correspond to their abilities, it may lead to automation misuse [58, 59], meaning errors and overall negative outcome. To that end, we believe ST can potentially help to calibrate the user's self-confidence by providing information about their peer's interactions with the system and the resulting outcomes.

Similarly to the work by [32] in this paper, we focus our efforts on qualitative evaluation of the ST framework but this time as applied to IT-security domain. At the same time, we incorporate quantitative measures to inform our qualitative results. The original study only measured the user's self-confidence. We extend the method by incorporating the trust measure. We are particularly interested in how ST affects trust and self-confidence, and as a result the user's reliance on the automated system.

3 STUDY METHODS

In this study, we aim to find out whether ST as a concept can be applied to the domain of system administration, and if sysadmins benefit from it. To answer our research questions, we conducted a mixed-methods study. Our approach was inspired by Ehsan et al. [32] and adopted their speculative design walk-through to the domain of network security and monitoring.

3.1 Speculative Design Scenario

In choosing and designing the task for the study, we consulted experts to provide their opinion on the most appropriate scenario as well as the insights that would make the speculative task believable and authentic for participants. When considering the daily workflows of system administrators, the choice of a prototypical task that all sysadmins would be familiar with proved to be challenging. System administrator is an umbrella term for a host of specialisations that presume different duties and types of expertise from the people performing the roles. As a result, system administrators' responsibilities may include monitoring and alerting, user permission and administration, software updates and installations among other things. The exact role and duties depend on the company's focus and the size of its computer network. The larger the network is, the more persons are involved in the task of system administration and the more specialised the roles become, whereby staff members divide the tasks. On the other hand, in smaller companies, it is not uncommon for a single person to be overseeing all responsibilities. The heterogeneity of the user group that we encountered, supports previous studies [10, 12, 43, 90]. For the sake of clarity and simplicity, in the context of our study, a sysadmin is a single person performing the task.

It is also worth noting that in search of a prototypical task, we focused on sysadmin workflows that would 1) benefit from AI-automation with real-life examples already being put into practice, and 2) would use AI as a decision-support mechanism requiring a human input to make the final decision. However, for the reasons mentioned earlier in the chapter we were not able to identify such a task. As result, we chose the anomaly detection scenario centred around an introduction of an AI-based intelligent firewall, as it represents the most common type of automation currently on the market. Our assumption was that most sysadmins would be familiar enough with the general goals, mechanisms and consequences behind the task, and with proper explanation would be able to follow the scenario without comprehension challenges. To ensure the participants' full understanding of the scenarios, we included a slide with the following background story before showing the mock-ups:

> *Imagine you are working at a medium-sized company as a system administrator and are responsible for overseeing the software of the computer system as well as the network infrastructure, including keeping the company's firewalls up-to-date and running. The company has recently made new acquisitions and expanded its operations. To help the system administrators inside the company, i.e. your colleagues and you, to run the network system smoothly, the company's executives introduced a new type of intelligent firewall solution. The new firewall uses Artificial Intelligence (AI) to detect anomalies in the network flow and alert you of potential threats and intrusion attempts in the system. It also suggests possible ways to resolve a conflict situation, which you can choose to either accept/allow or override/block. On one such occasion, the system detects an intrusion and proposes the following solution.*

Figure 1 shows the screenshots that were presented to the participants during the study. Before commencing the main experiment, we conducted a pilot study to test the comprehensiveness of the scenario and the overall procedure.

3.2 Study Procedure and Interview

We recruited 12 sysadmins serving as experts in our study. To ensure an appropriate fit and level of expertise, we followed the definition of sysadmins by Barrett et al. [10] to describe "those who use their technical, social, and organizational skills to architect, configure, administer, and maintain computer systems, including operating systems, networks, security systems, infrastructure, databases, web servers, and applications." This was done to recruit participants who performed these tasks as part of their job on a regular basis, regardless of their exact job title. We used social media networks, university mailing lists, industry forums, and personal communication to recruit the participants. The participants did not receive any compensation for taking part in the study.

Experiments were conducted via the *Zoom* [1] video conferencing software, lasted about 45 minutes and were video-recorded with the participants' consent for further analysis. Figure 2 shows the overview of the study procedure.

Through screen share the experimenter presented the scenarios via slides that depicted a mock-up interface as shown on Figure 1. We chose to include a static interface to eliminate any learning curve that a clickable prototype could potentially result in and to avoid any interference associated with it. Visual features were successively added or deducted during the experiments by changing

[1] https://explore.zoom.us/de/products/meetings/

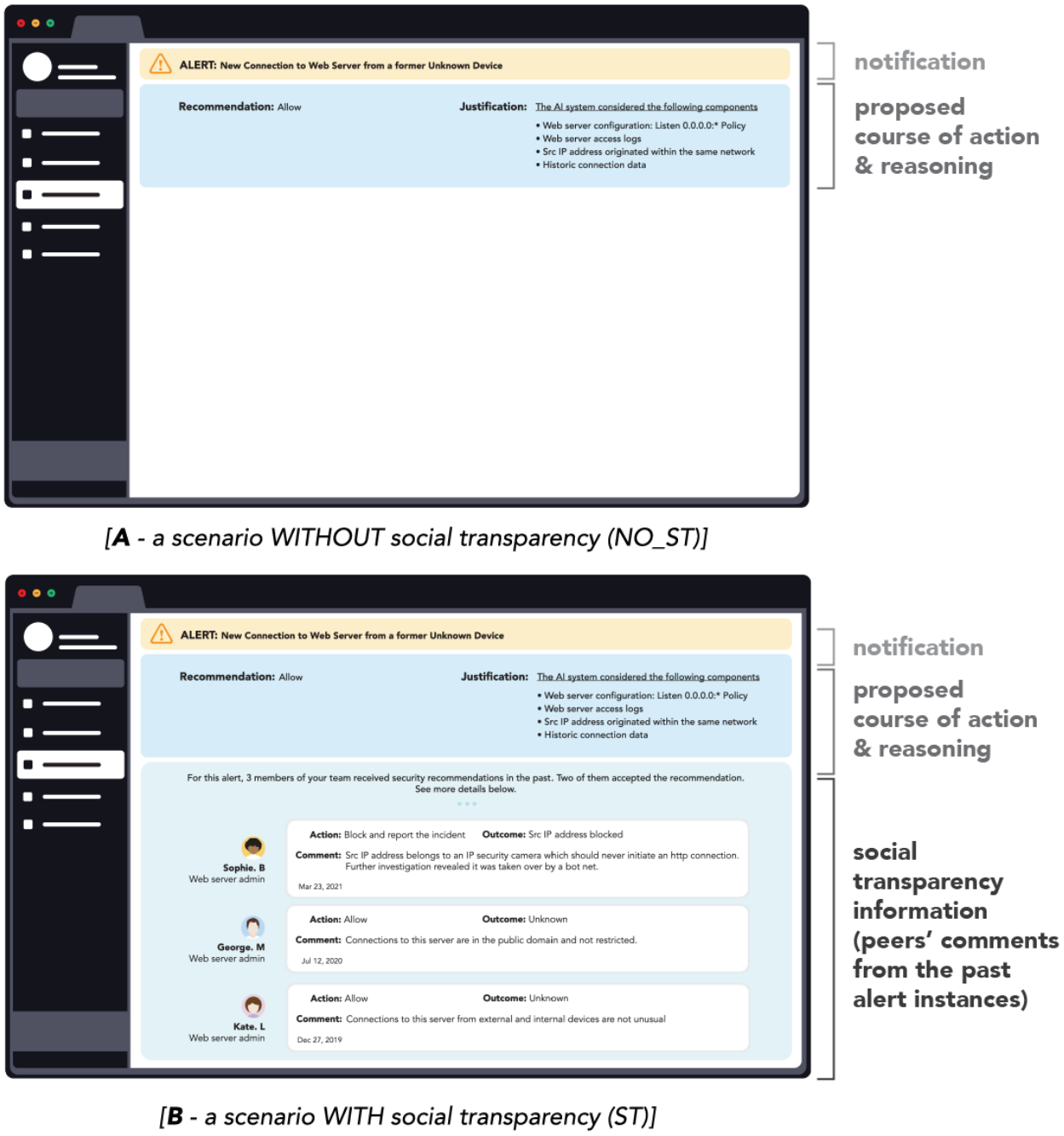

Figure 1: The design of the UI mock-up for the two scenarios, with and without ST.

the slides. The design of the mock-up closely followed the description in the study by Ehsan et al. [32] and the 4W approach *(Who? did What? When? and Why?)* - the constitutive element of ST as shown on Figure 3. *Who?* corresponds to the information about the person providing the comment, *What?* refers to the decision the person took, *When?* conveys the exact date of the interaction with the system, and *Why?* provides the reason for the person's choice of action. We used a within-subject design, and presented each participant with two consecutive scenarios.

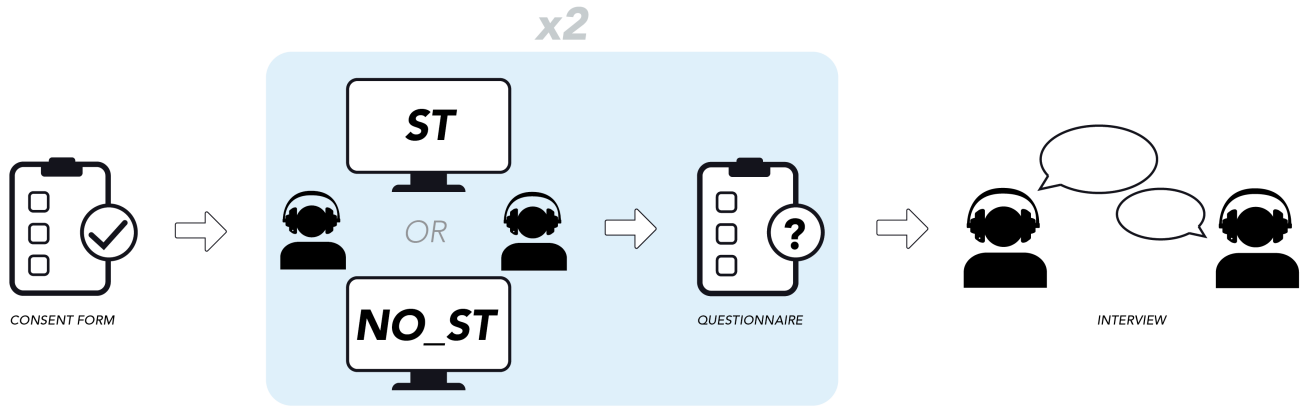

Figure 2: The overview of the study procedure. Before commencing the study, the participants received a consent form clarifying the details of the study and their rights. Upon signing the form, the respondents received the link to a Zoom call where they met the experimenter, and where the study took place. In the first part of the experiment, the experts filled out the demographics questionnaire and then were presented with two different scenarios, either with or without ST. They were encouraged to think aloud and verbalise their thoughts and opinions while confronted with the information they saw on the screen. After each scenario, the participants filled out a questionnaire. A semi-structured interview was conducted at the end of the study.

Using a fictive task, sysadmins were confronted with the scenarios where an AI-aided system detected an anomaly in the network and recommended a course of action.In both scenarios, the system displayed a "new connection" alert and proposed to continue allowing access from the web server to an unknown device. It also provided justification for its decision, i.e. information it took into account. However, in one of the scenarios (ST), the system also included information about the way the participant's fictional peers, other system administrators, in the company reacted in a similar context in the past (when AI made the same recommendation). The peers' comments in this case are an example of social transparency.

Having familiarized themselves with each situation, participants were asked to fill out a questionnaire that we describe in the subsection below. At the end of study, we conducted short semi-structured interviews conducted in English and German to ask participants about their thoughts on the ST framework and presented scenarios. We also inquired about the way the interviewees went about finding solutions to the problems they struggled with at work and the resources they tended to reach to the most.

3.3 Questionnaire

In our study, we presented a 3-part questionnaire to the participants. The demographics section collected the general information about the participants' age, gender, education, experience in network security and technical knowledge about AI. The second section contained three questions asking whether or not the participants trusted the systems' output based on the information provided, what kind of action they would like to take in response to this output and their confidence in their decision. We took the last two questions directly from the original study by Ehsan et al. [32] modifying them to suit the network monitoring context of the scenarios. The third section forms the trustworthiness part of our questionnaire. Its individual questions are inspired by the work of Ashoori and Weiz [8]. In line with their research, it covers the following facets, borrowing individual questions from their work: Overall trustworthiness, reliability, technical competence, and personal attachment [8]. Appendix A shows the detailed overview of the second and third sections of the questionnaire, which the participants filled out after familiarising themselves with each speculative scenario (twice in total).

3.4 Methodology for Qualitative Analysis

To analyse the qualitative part of our data, we used reflexive thematic analysis (TA) method by Braun and Clarke [14, 15] as described in [11]. We transcribed the recorded interviews using *otter.ai* [2] software and corrected the scripts afterwards to avoid transcription mistakes. Two interviews conducted in German were transcribed manually. The resulting scripts included think-aloud protocol and semi-structured interviews. As a starting point of the analysis process, two authors coded a representative sample of 10% of the data in an open-book coding approach using *Atlast TI* [3] software in line with Blandford et al. [11]. Through iterative discussions, an initial coding tree was established, which was then used by the two authors to code the remaining material. Any newly added codes were discussed in the process. We followed the six phases of thematic analysis as described by Braun and Clark [14, 15] to structure our analysis process. The same two coders were involved in the analysis to broaden the interpretation of the dataset. Through iterative discussions, we grouped codes into candidate themes, reviewed the emerging themes, and subsequently defined and named these themes. This resulted in four themes which are discussed in our findings.

[2]https://otter.ai/
[3]https://atlasti.com/de

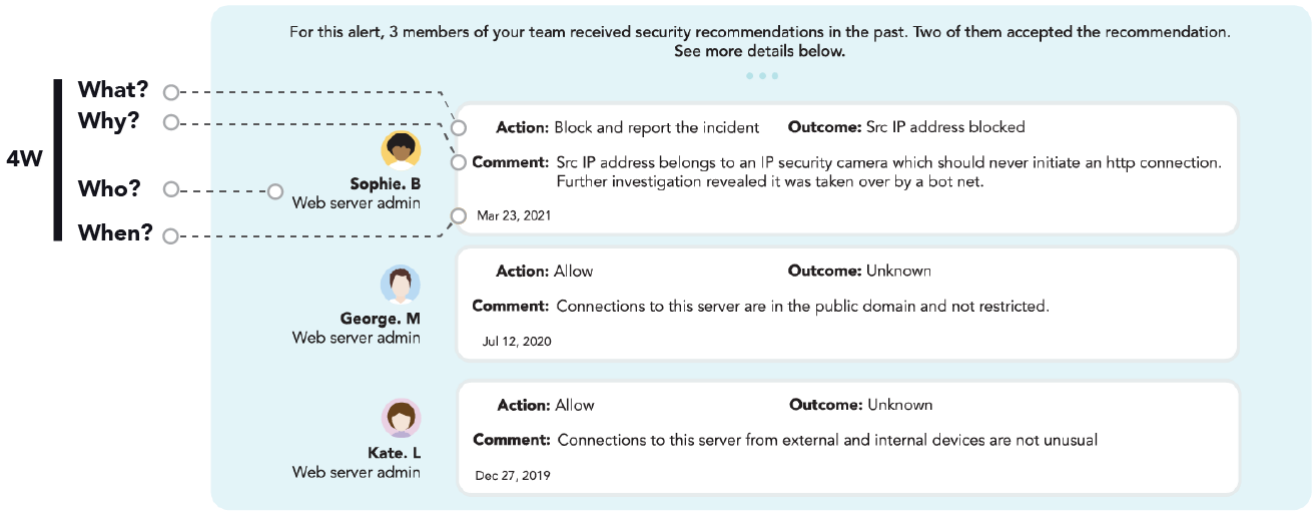

Figure 3: The 4W approach as applied to the design of the UI mock-up, based on the study by Ehsan et al. [32].

4 FINDINGS

Through the analysis of the interviews and the think-aloud protocols collected during the study, we developed four themes that describe sysadmins' user profiles, their existing workflows and challenges, their attitudes towards AI and opinions on the ST framework: 1) Sysadmins are sceptical towards AI, 2) Trust in the AI-based system must be earned, 3) Sysadmins need to understand the AI-based system, and 4) Sysadmins benefit from ST. Below, we describe our participant sample and discuss each theme in detail.

Participants. Overall, 12 experts participated in our study. Figure 4 shows a detailed overview of their profiles collected through a demographics questionnaire. All participants identified as male with youngest participant aged 27 and the oldest 59. The mean age in our study group was 36.6. Over a half of the respondents (8) reported to have received a Master's degree or higher. The experts current job description varied greatly and included the head of campus network, the head of IT Backend Services, IT managers and sysadmins to name a few. The average experience in the domain was 11.6 years, with minimum at 2 years, and maximum experience at 34 years. The experts rated their overall familiarity with network security 3.6 on average on a 5-point Likert scale from 'not at all familiar' to 'extremely familiar'. 10 out of 12 participants reported having no prior experience with (semi-)automated AI-based systems.

4.1 Sysadmins are sceptical towards AI

The most frequently recurring theme in our study is the sysadmin's general scepticism towards AI-based systems. Nearly all of our participants mentioned that they would not "blindly trust AI", while some were cautious to attribute too many potential benefits to AI:

> *"I wouldn't just blindly replace something with an AI-based thing. I've seen too many AI things out there in the last five years to be a little sceptical about that." (P6)*

This can be explained by a number of factors:

(1) **The nature of the sysadmin's training.** Concepts such as the principle of least privilege [51, 79, 80] and zero trust [19, 20, 34] contribute to the initial hesitation in adopting AI-based solutions in network monitoring and security, and make sysadmins an especially sceptical user group. The principle of least privilege is one of the fundamental access control postulates in information security where the users (human and programs) are granted only those access permissions they need to perform their tasks. Zero trust principle, on the other hand, is a concept that assumes that no user, network or activity is trustworthy to begin with, only when verified.

(2) **The feeling of responsibility.** Indeed, several participants explicitly mentioned feeling responsible for the system's decisions in case of a failure, especially in high-stakes scenarios. Most admitted that when uncertain about the right course of action, they would rather block a connection/access to a device than to risk an intrusion or an attack on the network:
"It's never a bad idea in practice to just be cautious." (P4)
It appears that 'better be safe than sorry' is a common policy among the admins.

(3) **Lack of (positive) experience with AI-based systems.** Some participants cited their knowledge about AI algorithms and expressed concerns that the system's decisions may not surpass that of a human user, or that the system might fail. It is however, possible that since most of our users have never encountered an AI-based system at their workplace their lack of experience could have an added influence on their sceptical attitude.

Scepticism aside, the experts we interviewed nonetheless confirmed the admins' need for automation at the workplace. As we mentioned earlier, one of the greatest challenges in

ID	Age	Gender	Level of education	Current role / job	Years of experience	Familiarity with network security*	Technical knowledge of AI**	Prior experience with AI-based (semi-) automated systems
P1	27	male	Master's degree or equivalent	Student / software engineer	5	4	2	not sure
P2	29	male	Master's degree or equivalent	IT project manager	2	4	2	no
P3	34	male	Master's degree or equivalent	System administrator	13	3	1	no
P4	31	male	Master's degree or equivalent	Research assistant	5	4	3	no
P5	59	male	High school diploma or equivalent	IT manager	34	3	1	no
P6	47	male	Master's degree or equivalent	Head of campus network	20	4	2	no
P7	34	male	Master's degree or equivalent	Research assistant	18	4	2	no
P8	58	male	Doctoral degree or equivalent	Head of IT backend services / deputy of CIO	21	3	1	no
P9	30	male	High school diploma or equivalent	Network specialist	4	4	1	no
P10	36	male	Trade / technical / vocational training or equivalent	System administrator	7	3	1	no
P11	23	male	Trade / technical / vocational training or equivalent	System administrator	3	3	2	yes
P12	32	male	Master's degree or equivalent	IT security specialist	7	5	3	no

** From 1 (not at all familiar) to 5 (extremely familiar)*

*** From 1 (not at all knowledgeable) to 5 (extremely knowledgeable)*

Figure 4: An overview of the expert participants' demographic data.

network security and monitoring today is the rapidly increasing volume of data and its complexity [13, 23, 75, 83]. It is difficult for the users to understand the incoming traffic and safeguard the network from malicious activities, as it requires time and significant human resources. With the exception of 2 participants, the experts we spoke to did not have any form of automation present at their workplace. The participants agreed that some level of automation would be welcome to help them to deal with repetitive basic tasks with low risk:

> *"Simplifying the process is always good, especially in a dynamic environment, like ours. Here, we have a lot of people coming and going, and people bringing their own devices, people changing devices. And I don't know what's going on! People create test beds, deconstruct test beds, reuse devices and so on. So it's always chaos, essentially. I would be really glad if some sort of system could automate at least the obvious choices. That would simplify my life significantly." (P4)*

4.2 Trust in the AI-based system must be earned

Connected to the previous theme is the idea of developing trust in the system (any AI system). Throughout the interviews, we observed that the participants were to a varying degree reluctant to trust an AI-based tool initially, but thought it possible to establish such a relationship. When asked about factors that could facilitate and accelerate the process, the vast majority of the experts reported that they would prefer to test a new system first to observe its performance over time:

> *"So, probably what I would do if I were to adopt such a system is to sort of let it run in parallel for quite some time before I start to blindly trust the system, let's say." (P4)*

Some mentioned that they would run a more established system they know well in parallel to compare the results, while the majority would spend some time verifying the system's decisions themselves to make sure it operated as intended. Naturally, consistency of the system's decisions and their quality seemed to heavily influence the admin's assessment. One of the less commonly cited, but interesting, factors was the similarity between the system's decisions and those of the user. It appears that if the system's decisions confirmed the user's own thinking, the user would trust the system more. This to some extent can be explained by the users' general tendency to be overly confident in their own abilities [37, 38, 59]. We see this as an important avenue for further investigation, as overconfidence in one's abilities can lead to errors and compromise the safety of the system. Colleague's recommendations would also reportedly influence the admin's decision to use the software, although the users would still prefer to decide on their own in the end.

Another commonly mentioned factor that would influence the development of trust in the AI-based tool was risk level. All experts agreed that in a high-stakes scenario where the gravity of consequences in case of failure is significant, they would be especially distrustful and cautious of the system's decisions and recommendations, and would want to have an input when choosing the course of action:

> *"So what degree of automated response is acceptable? What's at stake? Here somebody cannot access the website. Oh, my. And if it were like, okay, we would turn off the oxygen at the hospital, I would say no, no, let's have some human decision making here." (P6)*

On the other side of the spectrum are of course, low-risk scenarios where some participants admitted to be likely to rely on the system's decision for speed's and convenience's sake.

In addition to factors mentioned above, some experts we interviewed noted other aspects that could affect their judgement. Although cited less frequently, some respondents mentioned that the results of their personal research about the software would be a decisive factor, including technical information about the system, how well the software is established on the market and how reputable the manufacturer is. One of the participants stated the need for the system to provide feedback in case of failure to provide more transparent communication to the user.

4.3 Sysadmins need to understand the AI-based system

One of the central arguments of this paper was an explainable system as a prerequisite for secured network. Our interviews with the experts confirm that: The majority of the participants wished to "make an informed decision". The need for contextual information was the most commonly recurring topic and referred to two categories of information, the data and documentation that the admins normally have access to and the new information about the AI-based system. To be more precise, the admins expected to have access to

- Technical data about devices in the network,
- Connection history,
- And other relevant contextual information

regardless of the type of system at hand, AI-based or not. However, in the context of an AI-based tool, the experts stressed the importance of obtaining additional details, such as

- System specifications (f.ex. information about the AI algorithm and data it was trained on),
- Data parameters that influenced the system's outcome (including f.ex. confidence score),
- And information about the system's past performance (f.ex. past system logs and tickets).

Access to such information is meant to help the admin in a challenging situation, especially when they are unsure about the course of action suggested by the system.

However, more information doesn't equal better explainability or understanding, and our participants stressed the importance of having access to quality relevant information and wished for a more convenient way of gathering it. Several experts suggested a form of consolidated knowledge base where AI could help locate the necessary data:

> *"So essentially the tool has to provide convenience, right? It has to kind of be like a centralised knowledge space for you. Not just, you know, detecting something, but also providing you enough information so that you can make the right decision yourself essentially." (P7)*

In their current situation, the admins we interviewed reported reaching out to a broad range of resources, including the more unusual ones, like YouTube and Reddit. Google was the second most cited resource after colleagues. Among other resources that the respondents mentioned during the study were GitHub, forums, other experts, IT-security consulting firms, Google Scholar, resources collecting known vulnerabilities and the latest exploits, books and magazines, as well as product support.

Although the majority of the experts reported being satisfied or mostly satisfied with the resources they have, they also mentioned that in case of AI-based system, they would require additional information about the system's decision-making process. The majority showed opposition to AI based "black-box" models. As one of the respondents put it:

> *"To some degree, I'll have to be able to explain the decision that I can sign in the end." (P6)*

As stated earlier, several experts explained their need for such information by sense of responsibility and arising lack of trust when they could not know how the system arrived at its output, especially in high-stakes scenarios. We come to the conclusion that while the admin's current goal is primarily to understand their network, with the introduction of AI into their workflow, the users also need to understand the system in order to understand the network.

4.4 Sysadmins benefit from ST

When introduced to the ST framework in our study, the majority of participants reacted positively and thought that the concept, depending on its implementation, could provide benefits to the sysadmins. The experts found it especially helpful in situations when the information provided by the system was limited similarly to the no-ST scenario in our study. One of the participants noted:

> *"It was helpful for me, the information from the three persons. It helped me to make the decision." (P8)*

Several experts pointed out that ST can provide useful means of communication and passage of knowledge among colleagues. Especially in large companies with vast networks, it becomes difficult to know and to contact one's peers, let alone to have access to them when they leave the company:

> *"That's actually very important to know what other people have done, keeping track of that. And given a network of this size here, that would be very valuable, because a lot of that, you can't communicate through normal means." (P6)*

Our participants thought that having peer information in the system could make it more readily available and reduce time and effort when trying to contact and acquire additional information from peers.

It is worth noting that colleagues appear to be the most cited resource of external knowledge for the admins, although not just any colleagues. Indeed, we discovered that the mentions of peers by far surpassed those of online resources when talking about the resources that the participants tended to reach out to when unable to solve issues themselves. However, colleagues' reputation, expertise and experience played a significant role when it came down to asking for advice and trusting it:

> *"If I know that I have a colleague who is expert in this domain, then I'll ask them. If I know that, I have no person who has much experience in that, I won't ask a colleague." (P7)*

Interestingly, one of the participants mentioned that experience sometimes could be a disadvantage in situations when a more creative approach was required. They suggested that people with a lot of experience might be more "conservative" in their thinking.

Mirroring the scepticism towards their peers in real-life, the majority of the respondents voiced their concerns about the credibility and trustworthiness of the peers in ST context and as a result, helpfulness of their comments. Some drew parallels with online platforms and social media sites, where the user's credibility is hard to determine. To mitigate this issue, some suggested a rating or an up-vote system, similar to online platforms like Stack Overflow.

In addition to credibility, the study participants brought our attention to the need for certain quality standards for the peer comments to adhere to. First of all, the comments have to be relevant. One of the criticisms we received from the participants is the lack of clarity about what a 'similar' situation meant in the context of past peer comments, as implemented in our study. The experts wished for the comments that the system presents to be very closely related to the situation at hand and come from within the company. This means not simply past instances based on the same type of error, but based on the same IP address in question, for example:

> *"The comment would need to be associated closer to what the actual situation is, and not just say 'there was a device that wanted to connect, and I blocked it or not'. But like, maybe I can filter [the comments] by substring user agent, or whatever, and then see all the previous decisions that have been ever made for this kind of user agent string, or this IP range, or this scale IP location or whatever. So that I have more context and not just a giant list of things that I don't know how relevant they actually are." (P3)*

The experts also wished for the comments to be more "technical" or even have direct links to the tickets, logs and other relevant information. In general, technical information relevant for the current alert was valued more than peers' comments, and even in the scenario with ST, the participants wished for more information. This, however, can be explained by the design of the UI mock-up used in the study, which did not provide access to the resources that the admins might usually use.

Perhaps, the most surprising outcome of the expert interviews was the sentiment that ST could be potentially dangerous. Our participants explained it by their suspicion that some of their colleagues would follow the recommendation of another person without double checking the sensibility of their comments:

> *"When you see that they just allowed the traffic, then you instantly think 'well, it can't be that bad, right?'. So, pressing 'allow' is much faster, and I'm afraid that you don't really form your own decision anymore and don't look closely at the problem." (P2)*

They feared that especially in the cases when the user is closely familiar with the person providing the comment, they might overrely on their opinion and take the "if they say fine, it's fine" route. And if the colleague's past actions were wrong, the respondents feared that over-reliance could lead to a new mistake. Either way, ST can certainly influence the users decisions.

When it comes to potential benefits of ST, the experts mentioned a few other interesting aspects. One of the participants pointed out that the peer comments "made him stop and think", whereas without them he would have just allowed the connection. Another participant mentioned that ST comments could be used as an input for training the AI system further and improving its performance, and as a way of assessing its "learning" outcomes. One of the less expected benefits that the experts brought up was the possibility of using ST for team management. It was suggested that the peer comments could be useful in assessing the skills of the team members and "identifying biases", and therefore, signalling the need for additional training when the team's abilities didn't meet the requirements.

5 QUANTITATIVE RESULTS AND RECOMMENDATIONS

This section reports the results of the questionnaires that participants filled out during the study. All results, where possible, have been statistically analysed using *R* to investigate whether sysadmins changed their decision based on the ST information and how their self-confidence and trust were affected by the ST information. Effect sizes were calculated according to Robertson et al. [78]. It is important to note that many of the Likert scale results include ties, which makes the results of the Wilcoxon signed rank test with continuity correction less stable and can also result in an inflated V statistic which in turn inflates the effect size.

We compare the two experiment conditions in the following sections. We use the following abbreviations to differentiate them: Social Transparency (ST), describes answers to the questions that were given after seeing the scenario with the social transparency comments; Conversely, no Social Transparency (no-ST) refers to the answers to the questions that were given after seeing the scenario without the social transparency comments.

5.1 Impact of Social Transparency on Sysadmins Decision Making

Participants answered whether they trusted the system's output based on the information provided with either 'Yes', 'No', or 'Not sure'. For the no-ST group, 6 participants answered 'Yes', 3 'No', and 3 'Not sure'. The answers changed for the ST group showing 7 answers for 'Yes', 5 for 'No', and 0 for 'Not sure'.

After commenting on the trust, participants were asked to decide if they would follow the system's advice or override the recommendation. In the no-ST group, 3 participants would have followed the advice, while 9 would have overridden the recommendation. In the ST group, this changed to 7 participants following the advice, versus 5 overriding it.

When asked to rate their confidence in their decision on a five-point Likert scale, the no-ST group rated their confidence at a median of 3.5 ($mean = 2.83$, $sd = 1.34$). The rating improved for the ST group to a median of 4 ($mean = 3.58$, $sd = 1.16$). A Wilcoxon signed rank test with continuity correction found a significant difference between the two groups ($V = 28$, p-value = 0.0177, r = 0.532).

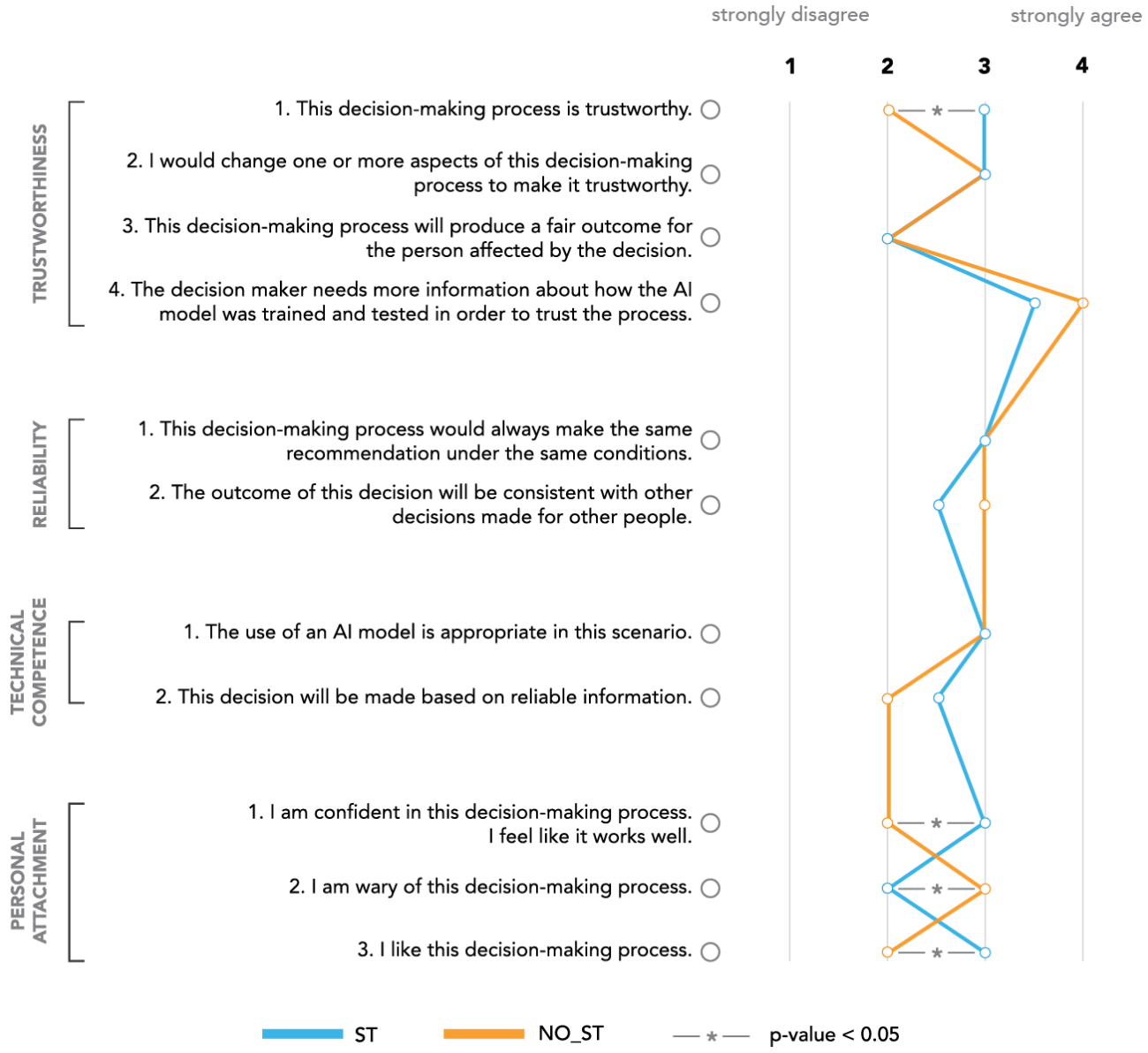

Figure 5: The median of all answers to the trustworthiness questionnaire by [8] for both scenarios.

5.2 Trustworthiness

The results of the trustworthiness questionnaire are presented in Figure 5 and Appendix B. This section reports on the four questions that showed significant differences in the answers of participants. The remaining seven questions did not show any significantly different answers. Participants in the ST group found the decision making process more trustworthy ($V = 32$, p-value $= 0.0418$, r $= 0.436$) rating it at a median of 3 ($mean = 2.67$, $sd = 0.89$) compared to the no-ST group rating it at a median of 2 ($mean = 2.08$, $sd = 0.67$). To the question whether they were confident in the decision-making process and felt that it worked well, participants in the ST group gave a higher rating at a median of 3 ($mean = 2.67$, $sd = 0.98$), rating it significantly higher ($V = 15$, p-value $= 0.0477$, r $= 0.454$) than the no-ST group who gave a median rating of 2 ($mean = 2.17$, $sd = 0.94$). Participants in the ST group were significantly less wary of the decision making process ($V = 0$, p-value $= 0.0147$, r $= 0.535$) with a median rating of 2 ($mean = 2.33$, $sd = 1.15$) compared to the no-ST group with a median rating of 3 ($mean = 3.08$, $sd = 0.79$). When asked if they liked this decision-making process, participants in the ST group gave a significantly higher rating ($V = 28$, p-value

= 0.0147, r = 0.535) at a median of 3 (*mean* = 3.08, *sd* = 0.79) compared to the no-ST group with a median rating of 2 (*mean* = 2.42, *sd* = 0.79).

5.3 Design Recommendations for ST in IT-security

Through combining the qualitative insights from the interviews and the quantitative results from our questionnaires, in this section, we derive five recommendations for improving the design of ST for the IT-security context. Generally, the experts reported to like and to be more confident during the decision-making in the ST scenario, and at the same time, appeared to also be less wary of it. We saw an increase in self-confidence and confidence in the the decision-making process after seeing ST information. This is in line with earlier works suggesting that access to more information results in higher self-confidence [59, 74]. In the case of our study this could be attributed either to the overall more information available to the users and/or to social validation and the sense of shared responsibility based on the peer's past decisions. While the respondents tended to consider the ST scenario more trustworthy, they reported the need to change a few aspects about the decision-making process to make it more thorough and trustworthy in both scenarios. Furthermore, during the interviews we learned that experts, although generally favourable towards ST information, wished for it to adhere to certain requirements to be useful. They emphasised that ST can be a valuable addition to their workflow but access to the relevant technical information is a priority. Finally, the experts' cautious attitude towards no-ST scenario can be explained by the overall lack of information about the system's background processes and the incident.

Based on our findings, we propose the following recommendations for incorporating ST into the sysadmin's workflow:

5.3.1 Peer comments need to adhere to a standardised structure. When providing information about the peers' past interactions with the system, it is crucial to have a standard for the structure and content of the comments. On a broader level, in addition to providing details about the past alert, the system's recommendation and the outcome of the incident, we suggest incorporating the 4W *(Who? did What? When? and Why?)* approach described by [32]. We advise then to include the following information:

- Description of the system alert and its recommendation,
- The person responsible for handling the incident - *Who?*,
- The final decision taken by the responsible person - *What?*,
- Date and time of the past incident - *When?*,
- Reasoning behind the responsible person's decision - *Why?*,
- The outcome of the decision if known.

5.3.2 Peer comments have to be relevant. Comments need to be closely connected to the current issue that the sysadmin is handling at the moment. While information on the same type of alert may be useful, information on the same device will be more actionable and should be favoured. The closer the parallel between the past and the present incidents is, the more likely the past solution is to fit the new problem. This will also make it easier for the sysadmin to find a better, more suitable solution in case a mistake was made previously.

5.3.3 Peer comments should link to past incidents. While adherence to a standard structure and relevance of the comments are crucial, it can be difficult to include every technical detail in a relatively short text block such as a peer comment. At the same time, the comment itself may contain a more subjective information such as personal interpretation or a guess that the admin should be able to verify or investigate. For that purpose, links to the data of the corresponding past incident should be included.

5.3.4 Peer comments should include people's qualifications. In other words, comments may provide details about the person behind the comment, such as their position in the company, expertise and experience. This may help the admin to assess the validity of the comment. At the same time, we suggest approaching this recommendation with caution, as this type of information may also bias some users and discourage them from judging the content of the comments more objectively.

5.3.5 Users should be able to provide feedback on peer comments. To some extent, feedback on the usefulness of a comment could help offset the bias mentioned in the previous suggestion associated with the peers' qualifications and experience. Even experts make mistakes. The feedback can then be added in a form of another comment, or through a rating system similarly to platforms like Stack Overflow, where the comments can be up-voted by the users.

With the list above we present five design recommendations for applying ST in IT-security . While this list is not exhaustive, it opens up a conversation on the role of ST in AI-based systems in the IT-security context. When properly designed and integrated, ST can be useful in facilitating explainability of AI-based systems introduced to the sysadmin's daily work. Especially when the the system's decisions are difficult to explain, as is usually the case with "black-box" models, ST could help sysadmins to understand its underlying mechanisms. This would provide the users knowledge about what the system can and cannot do, and help the admins to adjust their expectations about the software and assess the amount of input needed from their side. Being able to adequately evaluate the system's capabilities would leave less room for unwanted surprises, and the users would feel in greater control of the system and the network. ST has the potential to facilitate some of this control by encouraging the knowledge exchange among peers and integrate the human element into the security system.

6 LIMITATIONS AND FUTURE WORK

In this paper, we used a speculative scenario to glean more insights into the profile of our end users, sysadmins. At the same time, we measured reported trust and self-confidence of the participants in regards to a prototypical system taken out of the real-world context where external factors like time pressure could have had a significant influence on the participants' response.

When it comes to measuring trust, we focused our attention on the users' self-perception in our study, and did not consider other contributing factors that could have affected our results, such as accuracy of the system's suggestion. Trust as a concept is still subject of debate, even when it comes to the definition [36, 60], let alone ways of building, measuring and reporting it [17, 46, 52, 88, 91, 95].

To complicate the matter, it has been shown that trust is dynamic and can change over time [25, 45, 60]. Many researchers have been arguing in favour of shifting the focus from turn-based interactions with AI systems to continuous interactions [97] that are far more representative of the real-life situations and organisational structures, and allow to capture the user's behaviour more accurately. And although these are valid concerns, the effects of continuous interaction cannot be studied in a speculative scenario in a satisfactory manner, and are therefore, out of scope of this paper.

One of the avenues for further research is investigating the factors that could affect trust and reliance in AI-based systems with ST component, such as the quality of peer feedback, personal familiarity, or the aspects like professional standing of the peers and their experience. There already exist research on the ways the users determine a source credibility online and the effects of social engineering on their decision, especially in the context of social media [4]. It is possible, for example, that in a situation where the user is closely acquainted with their peers, their personal relationship could potentially cloud the judgement of their peers' comments and lead to undesirable consequences. Considering such factors and their influence on the effectiveness of the ST approach can prove instrumental in building secure and beneficial decision-support systems for sysadmins and other user groups.

7 CONCLUSION

This paper presents our investigation of system administrators as target users in the context of (semi-)automated AI-based systems. We examine the admins' profiles, their current workflows and attitudes towards AI-based systems and the challenges associated with their introduction into a traditional user workflow. We also test the applicability of the ST framework to the domain of network monitoring and security as a possible way of addressing some of the users' concerns, especially those pertaining to explainability, and appropriate trust and reliance in AI-based systems. Our results show that the ST framework can be applied to the IT-security domain and received favourable feedback from the group of experts we interviewed.

However, while ST has a great potential for assisting sysadmins in their daily tasks, this type of intervention alone cannot solve all trust and reliance issues in the IT-security domain associated with automation. While trust is a multifaceted issue and has many contributing aspects, the sysadmins' common sentiment is that AI has to earn the user's trust. Nonetheless, ST can be a step towards a more holistic approach to explainability in AI-based systems and a way to assist in knowledge acquisition and sharing among the peers. Furthermore, although not exhaustive on its own, ST can yield benefits in combination with other supporting factors to provide greater control over the system to the users and facilitate their decision-making processes by helping them to form a clear mental model of the system. As understanding not only the network but also the system becomes the primary goal of the users, we believe that ST can assist in introducing AI into the sysadmin's workflow. With this work, we contribute a first set of recommendations to facilitate this introduction. Still, more HCI research is needed in the domain of IT-security in this respect and we hope that our work will initiate further discussion on this topic.

ACKNOWLEDGMENTS

We would like to thank the Volkswagen Foundation and the Federal Ministry of Education and Research of Germany (BMBF) for sponsoring this work through the Wintermute project (award number 16KIS1127).

REFERENCES

[1] 2020. Vectra AI. Retrieved September 7, 2021 from https://www.vectra.ai/
[2] 2022. ExeonTrace. Retrieved September 7, 2021 from https://nextgen.exeon.com/
[3] 2022. Microsoft Azure Sentinel. Retrieved September 7, 2021 from https://azure.microsoft.com/de-de/services/microsoft-sentinel/#features
[4] Abdullah Algarni, Yue Xu, and Taizan Chan. 2016. Measuring Source Credibility of Social Engineering Attackers on Facebook. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*. 3686–3695. https://doi.org/10.1109/HICSS.2016.460
[5] Marco Angelini, Graziano Blasilli, Tiziana Catarci, Simone Lenti, and Giuseppe Santucci. 2019. Vulnus: Visual Vulnerability Analysis for Network Security. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 183–192. https://doi.org/10.1109/TVCG.2018.2865028
[6] Ariful Islam Anik and Andrea Bunt. 2021. Data-Centric Explanations: Explaining Training Data of Machine Learning Systems to Promote Transparency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 75, 13 pages. https://doi.org/10.1145/3411764.3445736
[7] Dustin Arendt, Dan Best, Russ Burtner, and Celeste Lyn Paul. 2016. CyberPetri at CDX 2016: Real-time network situation awareness. In *2016 IEEE Symposium on Visualization for Cyber Security (VizSec)*. 1–4. https://doi.org/10.1109/VIZSEC.2016.7739584
[8] Maryam Ashoori and Justin D. Weisz. 2019. In AI We Trust? Factors That Influence Trustworthiness of AI-infused Decision-Making Processes. *ArXiv* abs/1912.02675 (2019).
[9] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
[10] Rob Barrett, Eser Kandogan, Paul Maglio, Eben Haber, Leila Takayama, and Madhu Prabaker. 2004. Field studies of computer system administrators: Analysis of system management tools and practices. 388–395. https://doi.org/10.1145/1031607.1031672
[11] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI Research: Going Behind the Scenes. In *Synthesis Lectures on Human-Centered Informatics*.
[12] David Botta, Rodrigo Werlinger, André Gagné, Konstantin Beznosov, Lee Iverson, Sidney Fels, and Brian Fisher. 2007. Towards understanding IT security professionals and their tools. In *Proceedings of the 3rd symposium on Usable privacy and security (SOUPS '07)*. Association for Computing Machinery, New York, NY, USA, 100–111. https://doi.org/10.1145/1280680.1280693
[13] Raouf Boutaba, Mohammad A Salahuddin, Noura Limam, Sara Ayoubi, Nashid Shahriar, Felipe Estrada-Solano, and Oscar M Caicedo. 2018. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *Journal of Internet Services and Applications* 9, 1 (2018), 1–99.
[14] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3 (01 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa
[15] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* 18, 3 (2021), 328–352. https://doi.org/10.1080/14780887.2020.1769238 arXiv:https://doi.org/10.1080/14780887.2020.1769238
[16] Kirk Bresniker, Ada Gavrilovska, James Holt, Dejan Milojicic, and Trung Tran. 2019. Grand Challenge: Applying Artificial Intelligence and Machine Learning to Cybersecurity. *Computer* 52, 12 (2019), 45–52. https://doi.org/10.1109/MC.2019.2942584
[17] Matthew Brzowski and Dan Nathan-Roberts. 2019. Trust Measurement in Human–Automation Interaction: A Systematic Review. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, 1 (2019), 1595–1599. https://doi.org/10.1177/1071181319631462 arXiv:https://doi.org/10.1177/1071181319631462
[18] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (apr 2021), 21 pages. https://doi.org/10.1145/3449287

[19] Christoph Buck, Christian Olenberger, André Schweizer, Fabiane Völter, and Torsten Eymann. 2021. Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust. *Computers & Security* 110 (2021), 102436. https://doi.org/10.1016/j.cose.2021.102436

[20] Mark Campbell. 2020. Beyond Zero Trust: Trust Is a Vulnerability. *Computer* 53, 10 (2020), 110–113. https://doi.org/10.1109/MC.2020.3011081

[21] Bram C. M. Cappers and Jarke J. van Wijk. 2016. Understanding the context of network traffic alerts. In *2016 IEEE Symposium on Visualization for Cyber Security (VizSec)*. 1–8. https://doi.org/10.1109/VIZSEC.2016.7739579

[22] Leong Chan, Ian Morgan, Hayden Simon, Fares Alshabanat, Devin Ober, James Gentry, David Min, and Renzhi Cao. 2019. Survey of AI in Cybersecurity for Information Technology Management. In *2019 IEEE Technology Engineering Management Conference (TEMSCON)*. 1–8. https://doi.org/10.1109/TEMSCON.2019.8813605

[23] Checkpoint. 2020. *Security Report 2020 | Checkpoint*. Retrieved March 1, 2021 from https://www.checkpoint.com/downloads/resources/cyber-security-report-2020.pdf

[24] Sonia Chiasson, Robert Biddle, and Anil Somayaji. 2007. Even Experts Deserve Usable Security: Design guidelines for security management systems.

[25] Erin K. Chiou and John D. Lee. 2021. Trusting Automation: Designing for Responsivity and Resilience. *Human Factors* 0, 0 (04 2021), 00187208211009995. https://doi.org/10.1177/00187208211009995 arXiv:https://doi.org/10.1177/00187208211009995 PMID: 33906505.

[26] Cybereason. 2021. empow. Retrieved September 7, 2021 from https://empow.co/

[27] Webroot Smarter Cybersecurity. 2019. Knowledge Gaps: AI and machine learning in cybersecurity. Perspectives from U.S. and Japanese IT Professionals. (2019). Retrieved September 8, 2021 from https://www-cdn.webroot.com/6015/4999/4566/Webroot_AI_ML_Survey_US-2019.pdf

[28] Andreas Dieberger, Paul Dourish, Kristina Höök, Phillip Resnick, and Alan Wexelblat. 2000. Social Navigation: Techniques for Building More Usable Systems. *Interactions* 7, 6 (Nov. 2000), 36–45. https://doi.org/10.1145/352580.352587

[29] Constanze Dietrich, Katharina Krombholz, Kevin Borgolte, and Tobias Fiebig. 2018. Investigating System Operators' Perspective on Security Misconfigurations. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*. Association for Computing Machinery, New York, NY, USA, 1272–1289. https://doi.org/10.1145/3243734.3243794

[30] Selma Dilek, Hüseyin Cakır, and Mustafa Aydın. 2015. Applications of Artificial Intelligence Techniques to Combating Cyber Crimes: A Review. *International Journal of Artificial Intelligence & Applications* 6, 1 (Jan 2015), 21–39. https://doi.org/10.5121/ijaia.2015.6102

[31] Valentino Di Donato, Maurizio Patrignani, and Claudio Squarcella. 2016. NetFork: Mapping Time to Space in Network Visualization. *Proceedings of the International Working Conference on Advanced Visual Interfaces* (2016).

[32] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI Systems *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 82, 19 pages. https://doi.org/10.1145/3411764.3445188

[33] Upol Ehsan, Philipp Wintersberger, Q. Vera Liao, Martina Mara, Marc Streit, Sandra Wachter, Andreas Riener, and Mark O. Riedl. 2021. *Operationalizing Human-Centered Perspectives in Explainable AI*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3441342

[34] Dayna Eidle, Si Ya Ni, Casimer DeCusatis, and Anthony Sager. 2017. Autonomic security for zero trust networks. In *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. 288–293. https://doi.org/10.1109/UEMCON.2017.8249053

[35] Nick Feamster and Jennifer Rexford. 2017. Why (and how) networks should run themselves. *arXiv preprint arXiv:1710.11583* (2017).

[36] Andrea Ferrario, Michele Loi, and Eleonora Viganò. 2020. In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions. *Philosophy & Technology* 33 (09 2020). https://doi.org/10.1007/s13347-019-00378-3

[37] Baruch Fischhoff and Don MacGregor. 1982. Subjective confidence in forecasts. *Journal of Forecasting* 1, 2 (1982), 155–172. https://doi.org/10.1002/for.3980010203 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.3980010203

[38] Baruch Fischhoff, Paul Slovic, and Sarah Lichtenstein. 1977. Knowing with Certainty: The Appropriateness of Extreme Confidence. *Journal of Experimental Psychology* 3 (11 1977), 552–564. https://doi.org/10.1037//0096-1523.3.4.552

[39] Genua. 2021. cognitix Threat Defender. Retrieved September 7, 2021 from https://www.genua.de/it-sicherheitsloesungen/cognitix-threat-defender

[40] John Goodall, Eric Ragan, Chad Steed, Joel Reed, G. Richardson, Kelly Huffer, Robert Bridges, and Jason Laska. 2018. Situ: Identifying and Explaining Suspicious Behavior in Networks. *IEEE Transactions on Visualization and Computer Graphics* PP (08 2018), 1–1. https://doi.org/10.1109/TVCG.2018.2865029

[41] Robert Gove and Lauren Deason. 2018. Visualizing Automatically Detected Periodic Network Activity. In *2018 IEEE Symposium on Visualization for Cyber Security (VizSec)*. 1–8. https://doi.org/10.1109/VIZSEC.2018.8709177

[42] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* 51, 5, Article 93 (Aug. 2018), 42 pages. https://doi.org/10.1145/3236009

[43] Eben M Haber and Eser Kandogan. 2007. Security Administrators: A Breed Apart. (2007), 4.

[44] Josune Hernantes, Gorka Gallardo, and Nicolas Serrano. 2015. IT infrastructure-monitoring tools. *IEEE Software* 32, 4 (2015), 88–93.

[45] Kevin Anthony Hoff and Bashir Masooda. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors* 57, 3 (2015), 407–434. https://doi.org/10.1177/0018720814547570 arXiv:https://doi.org/10.1177/0018720814547570 PMID: 25875432.

[46] Robert R. Hoffman. 2017. *A Taxonomy of Emergent Trusting in the Human–Machine Relationship*. CRC Press, London. https://doi.org/10.1201/9781315572529

[47] Dennis G. Hrebec and Michael Stiber. 2001. A survey of system administrator mental models and situation awareness. In *Proceedings of the 2001 ACM SIGCPR conference on Computer personnel research (SIGCPR '01)*. Association for Computing Machinery, New York, NY, USA, 166–172. https://doi.org/10.1145/371209.371231

[48] Mikko Hypponen and Tomi Tuominen. 2017. F-Secure 2017 State of Cybersecurity report. *F-Secure, Tech. Rep* (2017). https://blog.f-secure.com/the-state-of-cyber-security-2017/

[49] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 624–635. https://doi.org/10.1145/3442188.3445923

[50] Alon Jacovi, Swabha Swayamdipta, Shauli Ravfogel, Yanai Elazar, Yejin Choi, and Yoav Goldberg. 2021. Contrastive Explanations for Model Interpretability. In *EMNLP*.

[51] Ira Ray Jenkins, Sergey Bratus, Sean Smith, and Maxwell Koo. 2018. Reinventing the Privilege Drop: How Principled Preservation of Programmer Intent Would Prevent Security Bugs. In *Proceedings of the 5th Annual Symposium and Bootcamp on Hot Topics in the Science of Security* (Raleigh, North Carolina) *(HoTSoS '18)*. Association for Computing Machinery, New York, NY, USA, Article 3, 9 pages. https://doi.org/10.1145/3190619.3190635

[52] Jiun-Yin Jian, Ann Bisantz, and Colin Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics* 4 (03 2000), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

[53] Faouzi Kamoun, Farkhund Iqbal, Mohamed Amir Esseghir, and Thar Baker. 2020. AI and machine learning: A mixed blessing for cybersecurity. In *2020 International Symposium on Networks, Computers and Communications (ISNCC)*. 1–7. https://doi.org/10.1109/ISNCC49221.2020.9297323

[54] Hyungseok Kim, Sukjun Ko, Dong Seong Kim, and Huy Kang Kim. 2017. Firewall ruleset visualization analysis tool based on segmentation. In *2017 IEEE Symposium on Visualization for Cyber Security (VizSec)*. 1–8. https://doi.org/10.1109/VIZSEC.2017.8062196

[55] René F. Kizilcec. 2016. *How Much Information? Effects of Transparency on Trust in an Algorithmic Interface*. Association for Computing Machinery, New York, NY, USA, 2390–2395. https://doi.org/10.1145/2858036.2858402

[56] Igor Kotenko, Igor Saenko, and Fadey Skorik. 2020. Intelligent support for network administrator decisions based on combined neural networks. In *13th International Conference on Security of Information and Networks*. 1–8.

[57] Sara Kraemer and Pascale Carayon. 2007. Human errors and violations in computer and information security: The viewpoint of network administrators and security specialists. *Applied ergonomics* 38, 2 (2007), 143–154.

[58] John Lee and Neville Moray. 1992. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35 (05 1992), 1243–1270. https://doi.org/10.1080/00140139208967392

[59] John D. Lee and Neville Moray. 1994. Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies* 40, 1 (1994), 153–184. https://doi.org/10.1006/ijhc.1994.1007

[60] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80.

[61] Philip A. Legg. 2015. Visualizing the insider threat: challenges and tools for identifying malicious user activity. In *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)*. 1–7. https://doi.org/10.1109/VIZSEC.2015.7312772

[62] Laetitia Leichtnam, Éric Totel, Nicolas Prigent, and Ludovic Mé. 2017. STARLORD: Linked security data exploration in a 3D graph. In *2017 IEEE Symposium on Visualization for Cyber Security (VizSec)*. 1–4. https://doi.org/10.1109/VIZSEC.2017.8062203

[63] Jiwei Li, Xinlei Chen, E. Hovy, and Dan Jurafsky. 2016. Visualizing and Understanding Neural Models in NLP. In *HLT-NAACL*.

[64] Qingzi Vera Liao, Dan Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020).

[65] Hong Liu, Chen Zhong, Awny Alnusair, and Sheikh Rabiul Islam. 2021. FAIXID: A Framework for Enhancing AI Explainability of Intrusion Detection Results Using Data Cleaning Techniques. *J. Netw. Syst. Manage.* 29, 4 (oct 2021), 30 pages. https://doi.org/10.1007/s10922-021-09606-8

[66] Roy A Maxion and Robert W Reeder. 2005. Improving user-interface dependability through mitigation of human error. *International Journal of human-computer studies* 63, 1-2 (2005), 25–50.

[67] Fintan Mcgee, Mohammad Ghoniem, Guy Melançon, and B. PINAUD. 2019. The State of the Art in Multilayer Network Visualization. (01 2019).

[68] Sean McKenna, Diane Staheli, C. Fulcher, and Miriah Meyer. 2016. BubbleNet: A Cyber Security Dashboard for Visualizing Patterns. *Comput. Graph. Forum* 35, 3 (jun 2016), 281–290.

[69] Sean Mckenna, Diane Staheli, and Miriah Meyer. 2015. Unlocking user-centered design methods for building cyber security visualizations. In *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)*. 1–8. https://doi.org/10.1109/VIZSEC.2015.7312771

[70] Tim Miller. 2019. "But Why?" Understanding Explainable Artificial Intelligence. *XRDS* 25, 3 (April 2019), 20–25. https://doi.org/10.1145/3313107

[71] Sandra R Murillo and J Alfredo Sánchez. 2014. Empowering interfaces for system administrators: Keeping the command line in mind when designing GUIs. In *Proceedings of the XV International Conference on Human Computer Interaction*. 1–4.

[72] Sandra R Murillo, J Alfredo Sánchez, and Enrique Sánchez-Lara. 2015. Enhancing Interfaces for Network Security Administrators with Legacy Attributes. In *Proceedings of the Latin American Conference on Human Computer Interaction*. 1–8.

[73] Milda Norkute, Nadja Herger, Leszek Michalak, Andrew Mulder, and Sally Gao. 2021. Towards Explainable AI: Assessing the Usefulness and Impact of Added Explainability Features in Legal Document Summarization. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 53, 7 pages. https://doi.org/10.1145/3411763.3443441

[74] Stuart Oskamp. 1965. Overconfidence in case-study judgments. *Journal of Consulting Psychology* 29, 3 (1965), 261–265. https://doi.org/10.1037/h0022125

[75] Mauro José Pappaterra and Francesco Flammini. 2019. A Review of Intelligent Cybersecurity with Bayesian Networks. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. 445–452. https://doi.org/10.1109/SMC.2019.8913864

[76] Zhuwei Qin, Fuxun Yu, Chenchen Liu, and Xiang Chen. 2018. How convolutional neural network see the world - A survey of convolutional neural network visualization methods. arXiv:1804.11191 [cs.CV]

[77] Marco Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. 97–101. https://doi.org/10.18653/v1/N16-3020

[78] Judy Robertson and Maurits Kaptein (Eds.). 2016. *Modern Statistical Methods for HCI*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-26633-6

[79] Jerome H. Saltzer and Michael D. Schroeder. 1975. The protection of information in computer systems. *Proc. IEEE* 63, 9 (1975), 1278–1308. https://doi.org/10.1109/PROC.1975.9939

[80] Matthew W. Sanders and Chuan Yue. 2018. Minimizing Privilege Assignment Errors in Cloud Services. In *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy* (Tempe, AZ, USA) *(CODASPY '18)*. Association for Computing Machinery, New York, NY, USA, 2–12. https://doi.org/10.1145/3176258.3176307

[81] Iqbal H. Sarker, Md. Hasan Furhad, and Raza Nowrozy. 2021. AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. *SN Comput. Sci.* 2 (2021), 173.

[82] Sofia A.M. Silveira, Luciana A.M. Zaina, Leobino N. Sampaio, and Fábio L. Verdi. 2022. On the evaluation of usability design guidelines for improving network monitoring tools interfaces. *Journal of Systems and Software* 187 (2022), 111223. https://doi.org/10.1016/j.jss.2022.111223

[83] Sophos. 2021. *Sophos 2021 Threat Report.* Retrieved March 1, 2021 from https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/sophos-2021-threat-report.pdf

[84] Tim Stevens. 2020. Knowledge in the grey zone: AI and cybersecurity. *Digital War* 1 (2020), 164–170. https://doi.org/10.1057/s42984-020-00007-w

[85] Cisco Systems. 2017. Cisco 2017 annual cybersecurity report. (2017). https://learningnetwork.cisco.com/s/article/cisco-2017-annual-cybersecurity-report-pdf

[86] Mateusz Szczepański, Michał Choraś, Marek Pawlicki, and Rafał Kozik. 2020. Achieving Explainability of Intrusion Detection System by Hybrid Oracle-Explainer Approach. In *2020 International Joint Conference on Neural Networks (IJCNN)*. 1–8. https://doi.org/10.1109/IJCNN48605.2020.9207199

[87] Christian Tiefenau, Maximilian Häring, Katharina Krombholz, and Emanuel von Zezschwitz. 2020. Security, Availability, and Multiple Information Sources: Exploring Update Behavior of System Administrators. 239–258. https://www.usenix.org/conference/soups2020/presentation/tiefenau

[88] Ehsan Toreini, Mhairi Aitken, Kovila Coopamootoo, Karen Elliott, Carlos Gonzalez Zelaya, and Aad van Moorsel. 2020. The Relationship between Trust in AI and Trustworthy Machine Learning Technologies *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 272–283. https://doi.org/10.1145/3351095.3372834

[89] Francesco Ventura, Tania Cerquitelli, and Francesco Giacalone. 2018. Black-Box Model Explained Through an Assessment of Its Interpretable Features. In *New Trends in Databases and Information Systems*, András Benczúr, Bernhard Thalheim, Tomáš Horváth, Silvia Chiusano, Tania Cerquitelli, Csaba Sidló, and Peter Z. Revesz (Eds.). Springer International Publishing, 138–149.

[90] Fábio Luciano Verdi, Hélio Tibagí de Oliveira, Leobino N Sampaio, and Luciana AM Zaina. 2020. Usability Matters: A Human–Computer Interaction Study on Network Management Tools. *IEEE Transactions on Network and Service Management* 17, 3 (2020), 1865–1878.

[91] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. 5, CSCW2, Article 327 (oct 2021), 39 pages. https://doi.org/10.1145/3476068

[92] Luca Viganò and Daniele Magazzeni. 2020. Explainable Security. *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (2020), 293–300.

[93] Artem Voronkov, Leonardo Horn Iwaya, Leonardo A. Martucci, and Stefan Lindskog. 2017. Systematic Literature Review on Usability of Firewall Configuration. *ACM Comput. Surv.* 50, 6, Article 87 (dec 2017), 35 pages. https://doi.org/10.1145/3130876

[94] Artem Voronkov, Leonardo A. Martucci, and Stefan Lindskog. 2019. System Administrators Prefer Command Line Interfaces, Don't They? An Exploratory Study of Firewall Interfaces. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 259–271. https://www.usenix.org/conference/soups2019/presentation/voronkov

[95] Jennifer Wang and Angela Moulden. 2021. AI Trust Score: A User-Centered Approach to Building, Designing, and Measuring the Success of Intelligent Workplace Features. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 54, 7 pages. https://doi.org/10.1145/3411763.3443452

[96] Alma Whitten and J. D. Tygar. 1999. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *8th USENIX Security Symposium (USENIX Security 99)*. USENIX Association, Washington, D.C. https://www.usenix.org/conference/8th-usenix-security-symposium/why-johnny-cant-encrypt-usability-evaluation-pgp-50

[97] Philipp Wintersberger, Niels van Berkel, Nadia Fereydooni, Benjamin Tag, Elena L. Glassman, Daniel Buschek, Ann Blandford, and Florian Michahelles. 2022. Designing for Continuous Interaction with Artificial Intelligence Systems *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3491101.3516409

[98] Tianyin Xu, Han Min Naing, Le Lu, and Yuanyuan Zhou. 2017. *How Do System Administrators Resolve Access-Denied Issues in the Real World?* Association for Computing Machinery, New York, NY, USA, 348–361. https://doi.org/10.1145/3025453.3025999

[99] Tetsuay Yamamura, Kouji Yata, Tetsujiro Yasushi, and Haruo Yamaguchi. 1989. A basic study on human error in communication network operation. In *1989 IEEE Global Telecommunications Conference and Exhibition 'Communications Technology for the 1990s and Beyond'*. IEEE, 795–800.

[100] X. Jessie Yang, Vaibhav V. Unhelkar, Kevin Li, and Julie A. Shah. 2017. Evaluating Effects of User Experience and System Transparency on Trust in Automation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 408–416.

[101] Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 295–305. https://doi.org/10.1145/3351095.3372852

[102] Fangfang Zhou, Wei Huang, Ying Zhao, Yang Shi, Xing Liang, and Xiaoping Fan. 2015. ENTVis: A Visual Analytic Tool for Entropy-Based Network Traffic Anomaly Detection. *IEEE Computer Graphics and Applications* 35, 6 (2015), 42–50. https://doi.org/10.1109/MCG.2015.97

A STUDY QUESTIONNAIRE (PARTS 2 AND 3)

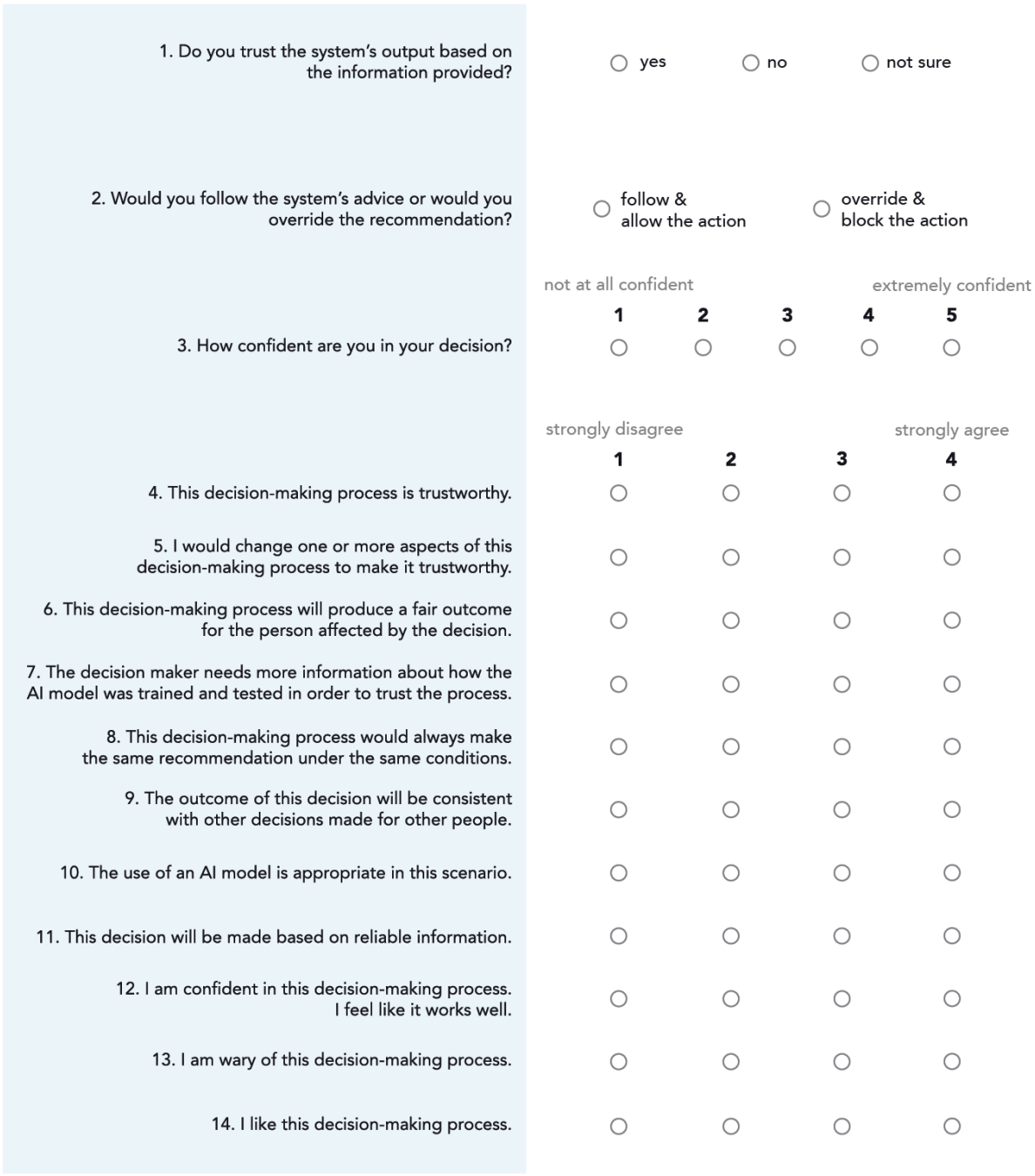

Figure 6: The questionnaire used in the study.

B RESULTS OF THE TRUSTWORTHINESS QUESTIONNAIRE

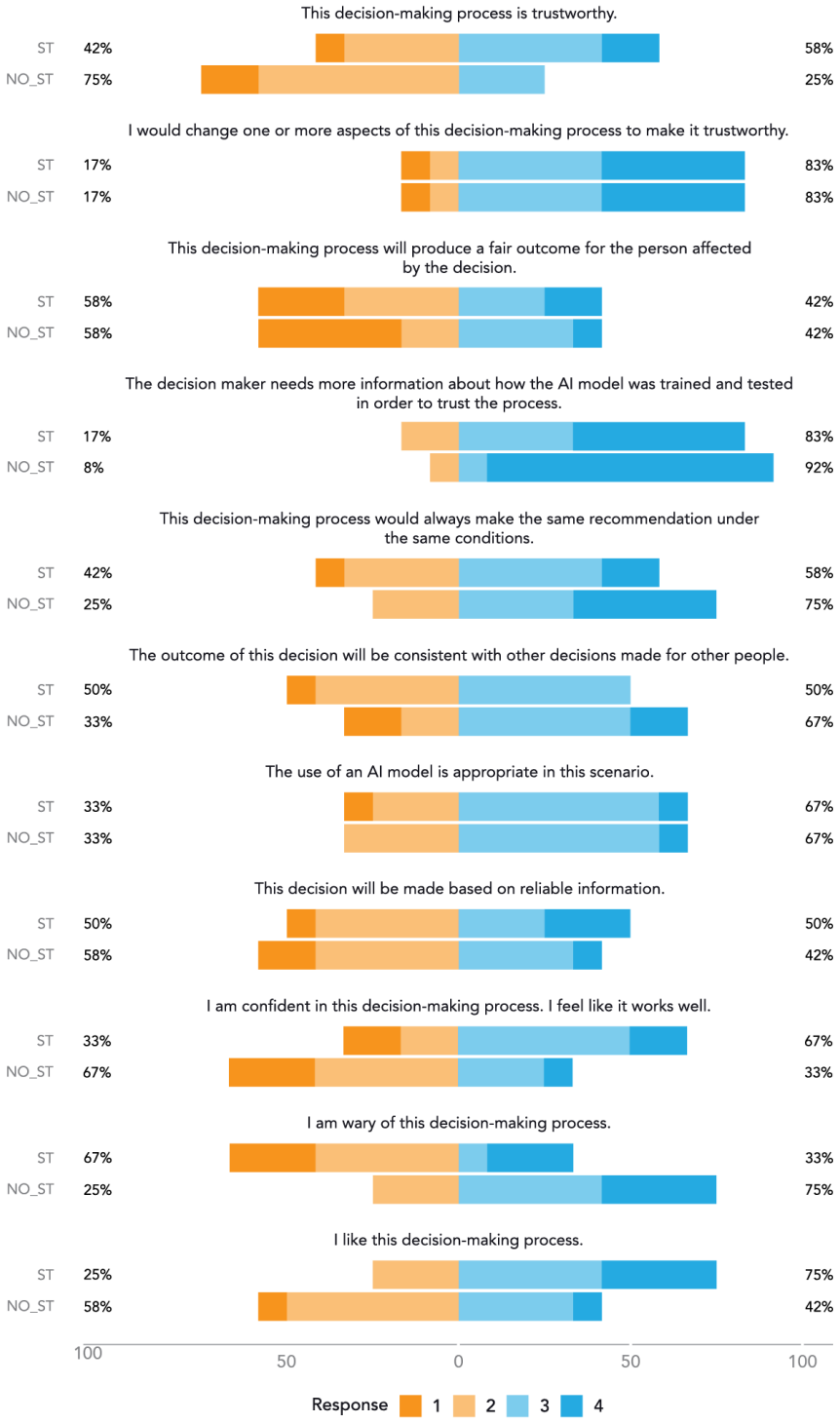

Figure 7: The participant responses to the trustworthiness questionnaire by [8].