Increasing Trust in a Decision Support System for Law Enforcement

A Human Computer Interaction Master Thesis Project
Utrecht University

Ive van Engelen
6196357

July 2024

Supervisor: Christof van Nimwegen
Second supervisor: Remco Veltkamp
External supervisor: Tamar Kastelein

Abstract

In this study, it was investigated how trust and transparency within a Decision Support System (DSS) can be enhanced to optimise decision-making and user acceptance. The research was conducted in the high-paced environment of the control rooms of the Dutch Police Force. Think-aloud sessions were organised with control room operators in order to gain insights into the feelings and opinions of the operators on two Explainable AI techniques: visual and textual explanation. These techniques were enhanced by several principles of Ecological Interface Design (EID). The analysis showed the emergence of five overarching themes: trust, time, convenience, integration and training. Trust was deemed the most important and controversial theme: against expectations, participants trusted the system's without the need for much clarification. This surfaced the concern that operators might trust the system too easily, and showed the importance of them using their own gut-feeling and experience, in combination with DSS, to come to the best decision. Further research can be done into the topic of explanations and trust, to further build on the results achieved by this study.

Contents

1 Introduction

1.1 Context

In our current day and age, the job of a police officer takes place both in the physical and digital world [81]. Society has become more and more digitalised [32]. Digitalisation comes down to the development that information and communication technologies (ICT) becoming increasingly prevalent in daily life [81]. This shift to digitalisation is not only caused by the changes in societal norms. It is also caused by the development of digital systems that can aid humans in their decision-making [8]. Systems have become increasingly complex, which has resulted in a change in the human role. Where humans functioned as the operators in previous times, this roles changed from having direct manual control to supervisory control [10]. Supervisory control entails that actions of automated devices are monitored by an operator. Thus, the human role changed from being in full control to supervising automated systems. Often, automation exists in a specific form, namely the form of automated decision support aids. However, this form of automation tells humans how to act instead of acting on their own [73]. When an automated decision support system is present, this tends to change the responsibilities and roles of the human operator. This is because the decision aid is the best cue for making decisions [60].

In the context of the police force, human operators play a significant role: for operational policing, the control room and its operators are crucial. The control room receives most of the emergency calls from citizens. After the call, these are prioritised and, if they have a high priority, immediately deployed to the units on the street. This way, the control room operators direct the operational police work. In addition, the control room dispatcher is approached and asked for advice by colleagues on the street. For example, if they want to obtain certain information, but also if central direction is needed. The larger the occurrence or incident, the more important and larger the role of the control room in that incident. In particular, an incident requires multidisciplinary cooperation. At all control rooms in the Netherlands police dispatchers work together with fire and ambulance dispatchers. In addition to coordinating an incident, the control room is also the life line for the officers on the street. In case of an emergency, an officer can use communication equipment, like C2000, to request back up. C2000 is a communication system through which police personnel can use to communicate by voice. All C2000 users are then immediately informed, so that the response speed is as high as possible. In addition to C2000, there are various other technical possibilities (such as a telephone) to relay or feedback reports and additional information.

High-impact incidents place many tasks on the dispatcher and time plays a significant role in completing these tasks. Commonly known is the term "the golden hour". Within that golden hour, the first 5 to 15 minutes are the most important and demand the most from the dispatcher. The dispatcher is directing, coordinating and supporting. In the first few minutes, an image is obtained, upon which a response is needed. This can be rapidly scaling up, as well as directing and coordinating units. The deployment depends on the type of incident and the information position. In addition to this there is high work and time pressure. That makes those first minutes crucial.

The IGMA-experiment explores whether AI can support the dispatcher and colleagues on the street in these first minutes after a report of an incident in which the suspect has fled and specifically while fleeing after a purgatory or ram-raid, which usually involves a different mode of escape (better thought-out escape due to advancing insight by the offender). So general escape via the highway (e.g. as fast as possible far away or across the border to Belgium or Germany) and specific escape methods of purgatories and ram-raids. The ultimate goal is to realise greater striking power as police. In order to increase this a Decision Support System (DSS) called IGMA (*Intelligente Geografische Meldkamer Assistent*) is developed. Development of the system started in the spring of 2021 and will continue into the foreseeable future.

1.2 IGMA

In the future, IGMA will be able to suggest the optimal deployment locations of the units involved. When the starting location of the fleeing suspect is known, it is simulated where it will move to in the future. These possible escape routes are based on all possible paths that a fugitive can take. However, it is taken into account that there is a higher probability of the fugitive taking the highway and remaining on these kinds of roads. Where other police units are located are also determined. Based on this data, the best location for the units to line up is calculated so that they can collectively wait to catch the fugitive with the highest possible chance of catching them. When new information about the fugitive is available, for example through an ANPR hit, the model can make another suggestion with the updated information. Automatic number-plate recognition (ANPR) is a technology that uses optical character recognition on images to read vehicle registration plates to create vehicle location data. When the fugitive is spotted, the pursuit can be initiated by the units and the fugitive can be contained. IGMA will make a deployment suggestion. The dispatcher may choose to forward this to the units on the street. These units are sent to the communicated spots via in-car navigation and/or on the phone. The use of IGMA can increase the chance of catching the fugitive, and lead to a more efficient deployment of police units.

In the IGMA project, state of the art AI will be deployed to, among other things, increase the catching power and police effectiveness during the golden hour. The focus of IGMA lies in the effective assignment of units. The goal here is to relieve the dispatcher by estimating what is possible based on the position of the unit as well as the route to the deployment location. This should save time regarding the positioning of units. Determining unit line-up locations is an important facet of the control room process, and IGMA aims to support the dispatcher to increase the catch rate in the golden hour.

1.2.1 Operations Centre

The moment a call is received, the first thing that is considered is which (available) unit is closest and is linked to the incident. If necessary, several units are linked. Depending on the incident, units will also be positioned. There is no standard, all-encompassing way to make a deployment proposal. Every incident requires a different approach and, as a dispatcher, you are always dealing with other ongoing incidents. Suppose there is a shooting incident in Harlem, several parties are immediately involved. There is also a direct communicative link with the Real Time Intelligence Centre (RTIC). Here, additional information is retrieved in available systems which can assist in catching the fugitive. If, for example, the registration number of the vehicle of the suspects is known, it is run through the ANPR. Should this produce a hit, a separate unit will drive on the ANPR hit. Based on the situation, the dispatcher estimates which unit should drive to which position. Colleagues on the street are listening and often anticipate the situation before they are called. Sometimes colleagues on the street do something other than what is instructed by the dispatcher since they have their own knowledge and experience they can build on. Often, they do this with the best intentions and based on their own estimations, but they are not always correct. Because the Operations Centre (OC) is understaffed and sometimes no position is assigned, officers will more often make their own estimates. Directing to positions cannot be done through a system: the dispatcher often mentions street names or hectometre poles. The colleagues on the street have to know this location or have to set up a navigation system to navigate to it.

The cooperation between the OC and the base teams (units on the street) are not evaluated. At the base teams, a debriefing is held, but the OC is not involved in this. The position of the OC has long been at the back of the line, while it should be the front of the line; they hold the final responsibility.

Evaluating the creation of deployment proposals does not happen. In fact, the entire work process is not evaluated, since there is not enough capacity in the control rooms. Calls cannot be re-listened. A switchboard operator can hardly have a normal break. In addition, dispatchers are in charge of peripheral issues that distract them from their core business. For example, they are responsible for arranging tow trucks and locksmiths when a door is broken down by the police. This leads to a busy dispatcher who is wasting time with arranging locksmiths, whcih results in greater waiting times for the 112 calls.

1.2.2 Why a DSS?

Within the Operations Centre (OC), several people are involved when a (major) incident happens. In the case of a ”flight” crime (such as an attempted robbery or liquidation) the role of the dispatcher is to ensure that units are placed in a position where the likelihood of an intervention is greatest. The placement of these units is primarily based on the ”gut feeling” of the dispatcher. Based on the experience of the dispatcher, the dispatcher’s level of knowledge regarding the crime and the local familiarity, he or she will direct and position the units.

Unfortunately, it appears that the "gut feeling" does not always lead to positive results. It greatly depends on the dispatcher what the outcome of the call will be. This is undesirable: in a context where people's lives are at stake, the best possible decisions should always be made. Decisions that are not influenced by an operators experience or feeling. This is where IGMA comes in. IGMA's goal is to suggest AI deployment locations, so that operators can make a more well-informed decision on where to place their units. IGMA can increase the consistency and the quality of the decisions made by the OCs.

Also, IGMA could improve the status of the OC within the base teams. IGMA could be used as a rationale why units on the street do have to listen to the direction. IGMA can support the dispatcher in making the best line-up location, so the base teams can trust the decisions of the operators more since they are backed up by a computer's decisions. The status of the OC will be further improved if IGMA actually delivers accurate results which would be better than the decisions made by only an operator.

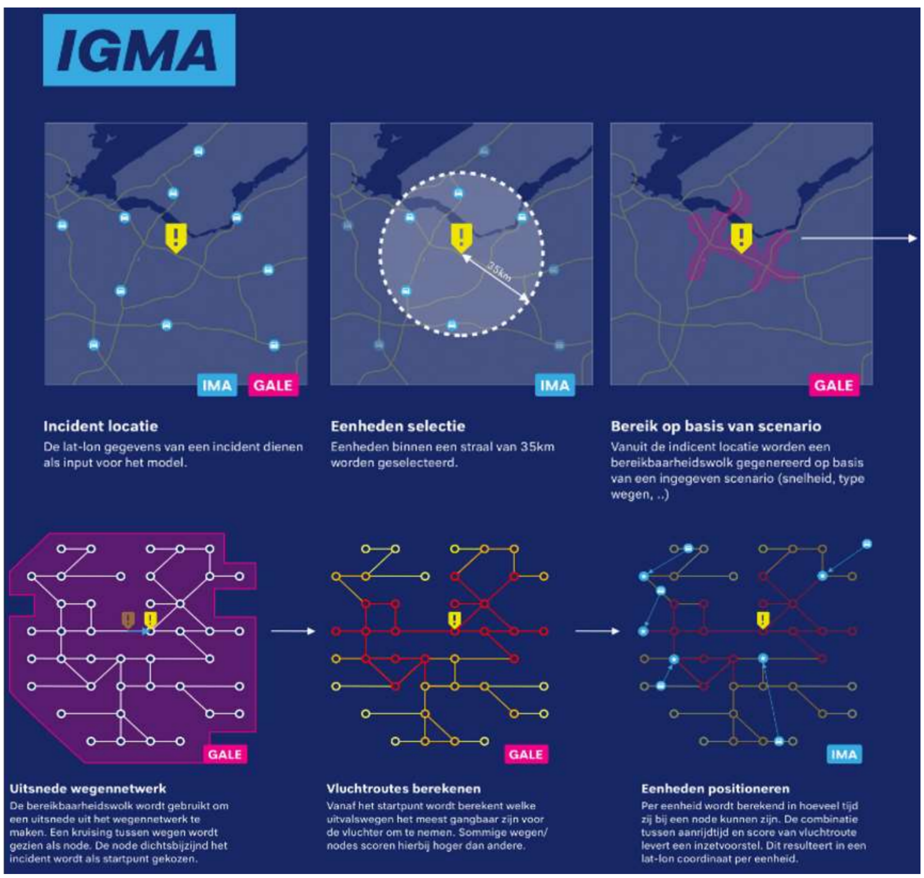

Figure 1: A Schematic Overview of IGMA (Dutch)

1.3 Problem and Challenges in the Criminal Circuit

1.3.1 Purgatories and Ram-raids

A large group of three to five hundred suspects from all over the Netherlands has been engaged in armed robberies for years. Most of these members come from a number of cities and neighbourhoods in the Randstad known to the police. Suspects seem to be getting younger and younger, much to the concern of the police. In February 2023, nine suspects were arrested. These individuals, among others residing in the Netherlands, are suspected of committing a series of ATM busts in Germany (believed to be 50 ATM busts). IGMA is an application that can help the police catch fleeing criminals. With the help of IGMA, the police can prepare faster and make smarter interventions on fleeing criminals.

1.3.2 Stolen vehicles

Vehicles are a major target for criminals. One of the biggest problems is theft. In addition to theft, other forms of crime occur, such as vehicle conversion and document forgery. Criminals are constantly finding new methods to innovate in this domain. Pooling information, adequate detection and better prevention all help to reduce the problem. In 2022, 5,973 passenger cars were stolen. This is an increase of 11% over the previous year. Of the stolen vehicles, c.a. 45% were recovered. Stolen vehicles are listed in the ANPR hit list. The chance of being caught is significantly increased by the positioning of units in case of an ANPR hit. Smart positioning and direction of units can contribute to the apprehension of one or more suspects of a criminal offence. This can take into account rules for escape behaviour of a specific group of suspects, thereby increasing the chance of being caught.

1.4 Current Situation Control Rooms

A baseline measurement has been conducted late 2023, before the IGMA experiment started. The experiment started when the Minimal Viable Product (MVP) was delivered in January 2024. During the first meetings with contacts in the "proeftuinen" (experimental havens) in November 2023, a demonstrator was used to show the ideas of IGMA. The purpose of the baseline measurement was to capture the current intake process within the control rooms, so that the progress and success of the IGMA tool can be measured and evaluated. In the IGMA experiment, the focus lies on one goal: selecting and positioning units on the street by the dispatcher.

Within the control room, various calls come in. These calls can range from (attempted) robbery to a fistfight, and from a neighbourly discussion to a homicide. For IGMA, the focus lies on escape incidents: calls where a fugitive is involved in a dead-to-rights situation, meaning that the fugitive can still be caught in that specific moment. Examples of types of incidents are a shooting or stabbing incident, kidnapping or assault involving fugitives, theft or robbery (burglary). An ANPR hit of a stolen vehicle falls into the category of escape incidents, as well. Several people are involved when a (major) incident happens in the control room. In an escape incident, the role of the dispatcher is to ensure that units are placed in a position where the likelihood of intervention is greatest.

In the control room, they currently use software called CityGis. This software allows them to see the live location of current units, the calls that came in and to which units the calls are connected. Currently, a model exists that can predict the range of a fleeing suspect. This model is called the Ring model. In figure 2, 3 and 4 it is shown how to activate this model. This model is currently not being used often by the operators. They indicate this is because it takes too much time to set up and adds too little value.

Figure 2: Step 1 of the Ring model

1.5 Existing Systems

A switchboard operator (dispatcher) works with 6 monitors, running many different systems at the same time. It is important to mention that the switchboard operator cannot operate additional systems. This means that IGMA must run within the systems already in use.

Since a wide variety of systems exist, it can sometimes be overwhelming for operators. They have to use all these systems at once, in a context which is stressful because of the lack of time. All the more important that IGMA runs within the systems that are already in use. The following systems are used within the control room:

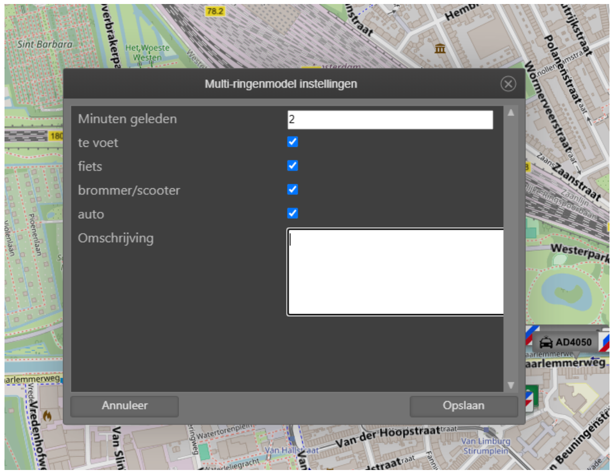

Figure 3: Step 2 of the Ring model

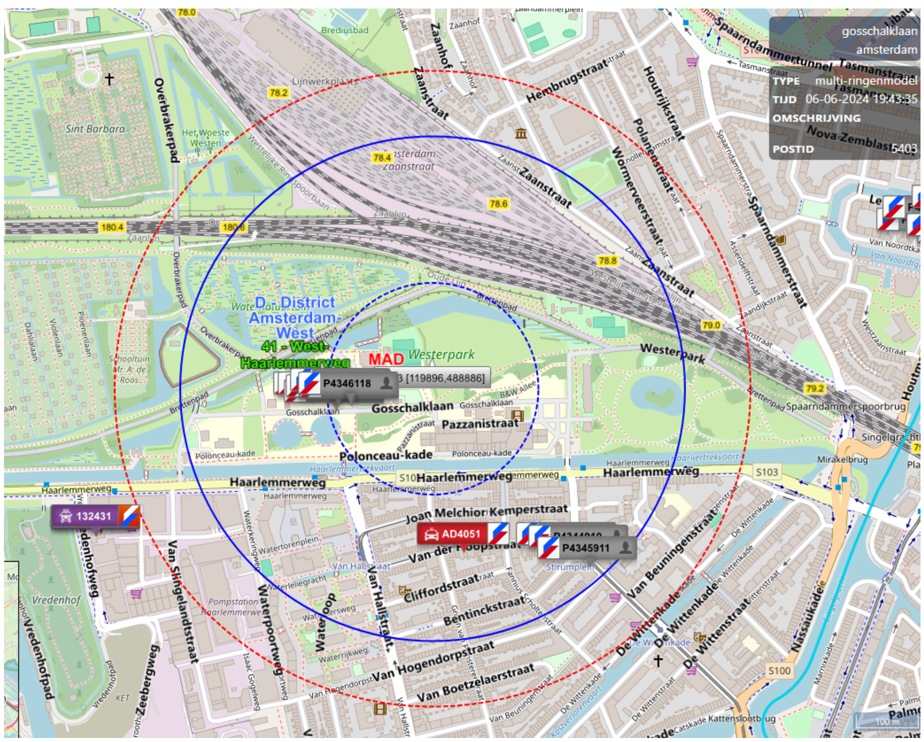

Figure 4: Final overview of the Ring model

1.5.1 OpGIS

Operational Gis: Geographic Information System. This system shows real-time information. Examples of information are where units, beacons and reports from GMS (Integrated Control Room System) are located. A possibility exists to use a reachability cloud. The disadvantage of this is, however, that it is not dynamic and only calculates the cloud from one (static) situation.

1.5.2 CityGIS

City: Geographic Information System. This system also works in real time. It provides the ability to make a deployment proposal. The software is managed by an external provider of software for the Control Room. This makes it difficult to make adjustments.

1.5.3 MS

MS is an Integrated Control Room System. Displays 112 incident reports coming in to the control room via an intake dispatcher. Also, wireless communication on the streets is facilitated by this system. Reports from citizens can be found here. Dispatchers are always dependent on the information they receive from the person reporting the incident. The type of incident is important and is rated on priority, ranging from 1 to 5.

1.5.4 C2000

C2000 is a communication system (walkie-talkie) to communicate by voice with colleagues in the car. Can also be used for group calls. The C2000 communication network is integrated into GMS.

1.5.5 PoltieAtlas

Displays the base teams on a map. There is a possibility to see where buildings are located, along with data about hectometre poles along the highway (for example). The system also contains aerial photos for all of the Netherlands, and these are updated regularly. A possibility exists to use a cloud which illustrates where available units can go in a given time. However, this cloud is static and is very slow due to retrieval of large amounts of data.

1.5.6 DragonForce

This software exists with the purpose of exchanging information. It holds a functionality that makes it possible to see where colleagues are. Also, with the software operators and policemen on the street can create a communication group with colleagues and exchange photos, documents, maps and deployment plans.

1.5.7 IRD

IRD is short for the Incident Response Dashboard. This dashboard unlocks real-time information from all kinds of police and open sources to enrich 112 reports. For example, from GMS, BVI-IB (Basisvoorziening Informatie, Integral Bevragen), Blueview, Twitter and Google. RTIC manages the information in IRD.

1.5.8 ANPR

ANPR is automatic license plate recognition. The abbreviation ANPR stands for the English term Automatic Number Plate Recognition. A special camera with ANPR technology reads license plates from passing vehicles. The read license plates are compared with a list of wanted license plates: license plates of drivers who have outstanding fines or who are wanted by the law, for example for an ongoing police investigation.

If a read license plate matches a license plate from the wanted list, police can take action. Whether and what action is taken depends on why the license plate was put into ANPR.

ANPR cameras are mainly located on highways. The cameras can be along the road or suspended above the road. Police officers can also carry such a camera in the car or place it in an inconspicuous spot along the road.

2 Research Questions

Previous research has shown that dependencies on DSS has been growing steadily over the last couple of years and will continue to do so in the future [11]. However, little research has been done about DSS in law enforcement. Moreover, cultural aspects such as trust and transparency have been left unexplored as well. Thus, the objective of this research is to explore this uncharted area of research and investigate cultural differences in the context of a DSS for law enforcement. This will be achieved by interviewing control room personnel and having them evaluate an updated version of IGMA that has been enriched with features that increase trust and transparency levels. The DSS that will be used to conduct this research is currently being developed by the Dutch local police. The main research question is thus as follows:

"How can trust and transparency within a criminal escape route visualisation and suggestion system for law enforcement authorities be enhanced to optimise decision-making and user acceptance?"

The sub-questions are formulated as follows:

1. What factors influence trust in a criminal escape route visualisation and suggestion system among law enforcement authorities?
2. How does the level of transparency in the decision-making process impact the perceived trustworthiness of the system?
3. What are the specific information and feedback mechanisms that enhance user understanding of the system's recommendations?

3 Literature Study

3.1 Decision Support Systems in Law Enforcement

A DSS can be defined as a computerised system, which supports and enhances the decision-making process in any given domain for any given decision maker [39]. Five categories of DSS exist: text-oriented, model-oriented, knowledge-oriented, database-oriented, communications-oriented [35]. In this specific context, a knowledge-oriented DSS is most interesting. This is a DSS that focuses on the knowledge and the development of knowledge. It can assist in developing knowledge, rules and strategies for solving problems [39]. DSS are applicable (and have been applied) in many fields, such as the medical field [26] and the financial field [83]. In the field of Law Enforcement, DSS have been shown to be successful at supporting the fight against identity theft for online business and consumers, as well as identifying people to support the investigation of crime [41].

Intelligent DSS have a similar purpose to traditional DSS, but differ in several aspects. One of these aspects is that intelligent DSS, when the focus of this system is to provide support for users, are enhanced with Artificial Intelligence (AI) technologies. The groundwork for both these technologies (AI and DSS) has not always been uniform. The aim of AI-related technologies might come down to the replacement of humans as the ultimate decision maker, at least for the long run. DSS' primary focus lies in support humans to make them the best decision-makers. Nevertheless, times have changed and so have the technologies. The technologies have advanced in such ways that the goals of both the technologies can be combined to be more beneficial to users in the end. These days, neural networks and algorithms are used frequently when looking at supporting technologies for DSS. When intelligent DSS were in an early phase of development, rule-based systems were often the technologies that supported them [3]. Many application fields for intelligent DSS exist. An example is the automatic analysis and allocation of images. Another example is expert systems. These systems are adaptive computer programs that can develop novel knowledge. They do this in particular fields, supported by reasoning mechanisms [87].

Intelligent DSS are designed for multiple purposes. One of these purposes is to adapt to user's processes, such as problem-solving and behaviour. They do this by using mechanisms similar to machine learning to align with the preferences, wants and needs of the users. Because of this, research into this topic has increased, along with their usage in practice. The users needs are constantly being tailored to and as a result, users lose a certain amount of control, as well as the fact that the system decides when and how to adapt, often without consulting its user first [33]. As a result, the behaviour of intelligent DSS becomes more challenging to predict because the constant adaptations lead to different outputs, even when the input for the system is unchanging. More often than not, IT background processes are not transparent and comprehensible for most users. In the case of intelligent DSS, the difficulty for users to understand their functioning becomes even greater [56] [76]. Consequently, it is of great importance for users to build and increase their trust in intelligent DSS, even though the comprehensive power of these users is low and their uncertainty is high.

Law Enforcement organisations face challenges similar to these, but also face a number of other challenges when implementing DSS. One of these challenges is to handle large quantities of data [39]. These enormous volumes of information and records that are linked to crime and criminals are difficult to handle. An intelligent support systems such as DSS can provide helpful assistance in criminal investigation, detection and other facets of the Law Enforcement environment. A DSS can play a role in the facilitation of recordings, retrieving certain analyses and sharing of important information [27].

Decision making is a process that depends on many factors. Factors include context, the way a person perceives and understand cues and what the personal values of the decision maker are [72]. Two factors stand are of great importance on decision making: cognitive perception and values. These factors influence how certain conditions are interpreted and responded to by a decision maker [2] [57]. These factors are important to take into consideration. Since the central and thus crucial part of a DSS is the interaction with humans, research must be done to investigate the cognitive perception and values of the humans it is being developed for. If that is the case, a successful DSS can be implemented that works seamlessly with their human counterparts and achieve the best possible results [80].

3.2 Trust in DSS

First and foremost, what is trust? And what does it entail? Four approaches are acknowledged about the definition of trust [23] [52]: it is a belief or a collection of multiple beliefs [9], emotional feelings [38], and intention [50], and a mixture of these fundamentals [52]. In this study, the main focus will be on trusting beliefs. This is because these beliefs have been identified as having a large impact on trusting intentions [44] [52]. According to Wang (2008), "trusting beliefs concerning a decision support technology include one's perceptions about the competence, benevolence and integrity of the technology". When and how trust is formed is very complex. Previous research has shown that trust can be formed because of multiple reasons, such as reasons based on institutions, knowledge and calculatives [18] [53]. Since people tend to treat computers as if they are humans (or social actors), humans tend to respond to computers based on the rules that are applied in the human world (like in human relationships) [70]. Similar to relationships, trust evolves with interactions over time [44]. A high level of trust is necessary when creating systems that aid humans in their daily tasks; systems similar to DSS. When a certain level of trust is missing, users cannot utilise the system to its full potential [4].

Other studies have shown great influence of reliability on trust as well. Research from Sanchez et al. (2004) states that reliability plays an important role in the process of forming trust, among reliability and reliance. Their results showed a significant decrease in reliance and trust when the reliability of the automated system was changed from 100% to 80% [73]. They also show that age can influence the perceived trust as well. The negative effects of unreliable automation were larger in older adults, which can greatly influence their ability to multi-task in high-tech environments [73].

In research regarding Information Systems, trust is a concept that is studied as a belief. This belief is being described as a concept based on three factors: benevolence, integrity and competence [95]. The transfer of these concepts to the world of IT systems is represented in different settings, for example in the setting of recommender system adoption. Originally, this transfer was used to elaborate trust in interpersonal relationships but has proven to be insightful in the IT domain as well [38] [89]. More recently, the research into the Information System domain has developed towards the evaluation of considerations in reliable and responsible system adoption [90]. Similarities exist between the growing automation of IT and the emphasis that is being placed on the interaction between people and systems. Research has been done in other domains, for example ergonomics and human factors, that investigated the relationship between trust and growing automation, as well [43]. This is the reason why some researchers promote focusing on human trust in the artefact itself, instead of evaluating trust in AT based exclusively on findings on interpersonal trust [42] [54] [79]. Lankton et al. (2015) prove that when selecting certain conceptualisations of trust, the level of anthropomorphism of the exact IT artefact should be taken into consideration. They state that the more "human" a system appears to be, the more suitable the measurement of trust is. This measurement is based on the earlier mentioned concepts of benevolence, integrity and competence.

On the contrary, interpretations of trust as discussed above seem to be less appropriate when the level of automation increases and the IT artefacts appear to be less human-like. When responses of Conversational Agents (CA) fail, it negatively influences the users' view of this CA in several facets, such as service contentment, belonging and perceived humanness [16]. Nonetheless, in increase in trust in specific behaviours of a CA can be measured when a proper context is provided by the CA why a certain mistake was made [63]. Whatever the perspective, three primary groups of trust determinants have surfaced in the domain of DSS [88] [95]. Both person-related factors and situation-related factors exist. Person-related factors encompass perceptions, attitudes and characteristics of a user. Situation-related factors represent external factors that impact the trust relationship dynamic users and DSS have. Nevertheless, the most notable factors are system features. These features mostly have to do with the performance of the system [28]. More often than not, the most important thing is the reliability of a DSS. When the output of the system is perceived by users as being consistent, the more likely the users are to build trust with the system [86]. On the contrary, unreliable behaviour of a DSS have a negative impact on the level of trust the users have. For example, a total system failure can have a negative impact on trust [55] and other malfunctions such as inaccurate results or false alarms can result in users losing trust, as well [17].

Research has also shown that mechanisms related to feedback and transparency have a great effect on trust in DSS [14] [66]. Mechanisms comparable to these can justify failures and explain the processes that underlie decisions to users, so that they can evaluate the reliability of the DSS. This can result in the prevention of user resistance [47] [91]. Many different works recognise the influence and importance of system-related failures on the trust relationship [48] [96].

3.3 Transparency

By making the decision-making processes in DSS more transparent, user's trust in the system can be established [78]. However, an overload of transparency can have contrary effects; with too much transparency, user trust can decrease [37]. For DSS that are not 100% reliable, and have a tendency to make mistakes every now and then, transparency can be especially useful. When being transparent about failures and revealing how and why they happened, the loss of trust can be mitigated when errors and failures happen[49] [33] [47]. This is why the main concern when designing DSS is that results can be understood by users, next to the fact that the output the system gives is accurate. Both correct and incorrect results should be understood by users so they can better substantiate their decisions [24]. Therefore, it is necessary for a transparent representation of system's output that a good match between the perceptions of users and system properties exists [14] [66].

Several competent ways exist for creating a more transparent DSS. One of these ways are explanations [67]. They facilitate a better understanding in users; the how and why of a DSS becomes much more clear when it is explained how certain decisions and behaviour emerge [29]. Besides, explanations can increase both the confidence in the behaviour and the output of intelligent systems [51] [93]. When using explanations, it is of great importance that the form and content are accurate. This will make the explanations more effective [29]. Recently, research has been focused on visualisation as well. Certain techniques can be applied to boost the transparency of a complex algorithmic process. This increased the level of trust [36].

Gönül et al. (2006) discriminated between short and long text-based explanation when evaluating the different structures and features of explanations. They did this in combination with high and low suggested confidence in the results. Statements with high confidence and long explanation had proven to be the most effective type of explanation in respect to the user acceptance of DSS [24]. Kizilcec (2016) concluded the need for balance in the level of transparency. When the amount of information is too big, doubt and distrust are increased [37]. Schmidt et al. (2020) that a level of transparency that is too high, can have unwanted effects for providers. Moreover, the original vision for a DSS is to reduce cognitive burdens, and with an overload of information, users would experience just that. However, when a system is overly simplified the lack of information may also have a negative impact on the trust relationship and lead to distrust [40].

Literature has shown that certain levels of transparency have a positive influence on trust in DSS. However, it is unclear how transparency measures are to be structured, especially when looking at the content and timing of the measures. This is particularly prevalent when erroneous decisions have already impacted the trust the users has in a DSS that has not been built yet [36].

3.4 Explainable AI

The concept of Explainable AI has been around for a while. Literature on this topic dates back to the seventies and eighties of the twentieth century [77] [82]. In this literature, the authors speak about knowledgeable systems that explain results via rules that have been applied to them. Since then, a lot of research has been done about AI and its context. Since it began, research have spoken about the explanation of AI results by intelligent systems, particularly when it comes to decisions. For example, when a rule-based intelligent system rejects a certain credit card payment, it is necessary for this system to explain its arguments for choosing this negative decision. All of the rules and knowledge an intelligent systems is built upon, are defined by humans, human experts in most cases. Since this is the case, the rules and knowledge are easy for actual humans to explain, accept and understand. A decision tree is an example of a method with explainable structure. A decision tree can be found in Figure 5, which illustrates a solution path in the decision tree. This tree presents the reasoning of a final decision

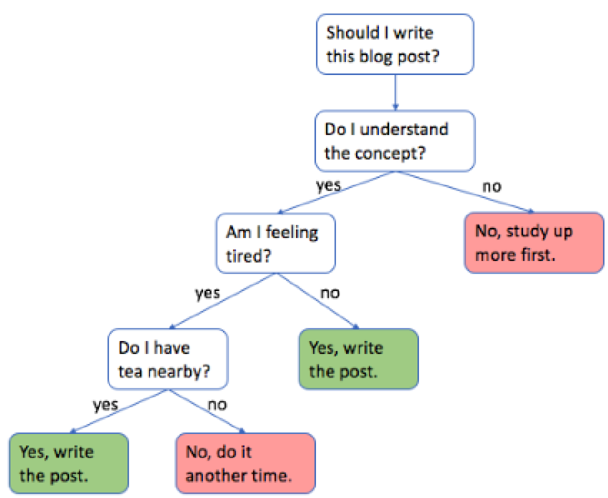

Figure 5: Decision Tree [94]

Recently, the aim of researchers in the field of AI is to transform the black-box that is neural networks into transparent systems. As is presented in figure 6, two central strands of work exist in Explainable AI (XAI): transparency design and post-hoc evaluation. The transparency design explains how the model functions, according to the developers. Furthermore, the transparency design tries to do three things: (1) understand the structure of the model, (2) understand the single components, and (3) understand the algorithm's training [94]. The post-hoc explains more about the results; how certain results came to be and what exactly they are about. This is done in the view of the users. It tries to do three other things, as well: (1) give analytic statements, so for example why certain products are recommended in a webshop, (2) give visualisations, like classifying pixel importance in object classification, and (3) give explanations by giving examples [94].

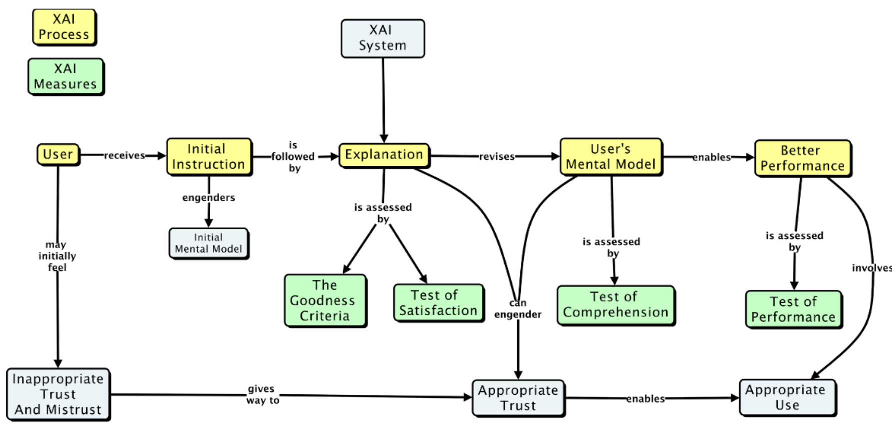

Figure 6: Explainable AI [94]

Attention to AI has been great from the start, but has recently increased around the world, mostly in the context of industries and research. An "Explainable AI program" was funded by DARPA in April of 2017 to improve the explainability of the decisions made by AI [15]. Other examples of big (government) projects to support the development of XAI exist, as well [88] [25]. The importance of XAI is ever-increasing to all groups of stakeholders. These groups include, but are not limited to the developers of the systems, people that are affected by these systems and the users themselves. XAI is especially important to the people who have to use the systems in their daily lives, either for professional or personal reasons. A decision maker, when using intelligent systems, needs to understand the reason the systems recommends/makes a decision. A flight-control operator needs to understand why their systems would recommend sending plane A to runway 3 instead of runway 2 when there is a storm coming, for example. This gives the operator insights into the reasoning of the system, which results in making a more deliberate decision. It is of great importance to certify the AI inference works as expected; otherwise high costs and even dangerous situations can emerge [94]. In a study done by Caruna et al. (2015) [13], this point is illustrated. They presented an AI system that was trained for the cause of predicting pneumonia risk. In this case, the systems had come to the wrong conclusion about a patient because the training data had been biased. This resulted in a diagnosis that was incorrect, which made the treatment dangerous and put the patient at risk. This example illustrates how important it is for a system to be easily understandable and transparent; if the doctors knew on which data the diagnoses was based, they could have easily spotted the bias in this data. These kinds of insights will remain undiscovered if certain intelligent systems remain black-box-like. This concept is illustrated in figure 7. This figure shows the input that is given to the system, which is a black box (in this case), and is then translated to output to the users. When one brings transparency into the mix, the black box will output its results directly to the user which removes the black-box effect altogether: the user understands what is happening in the black-box, thus removing its confusing nature.

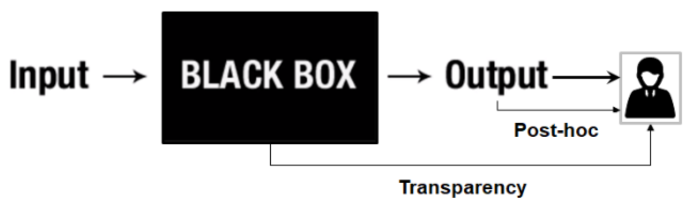

Figure 7: The Black Box of AI [94]

Obviously, in XAI, explanations play an important role. An explanation is not an attribution of statements; rather, it is an interaction. What can be seen as an explanation namely depends on several factors: what the needs of the user are, what the background of the user is (previous knowledge/skills) and most importantly, what the goal of the user is [32]. These statements lead to the examination of an AI system's context, before it is developed. This context can be expressed in questions that arise for users. These questions can be considered triggers. In other words, when a user seeks an explanation, it is actually a need to satisfy their purposes or goals [32].

Previous research show assertions about what a good explanation actually is. An accord exists on the fact that clarity and precision play a significant role. This way, explanations can be judged and qualified as being good or not [32]. Explanation Satisfaction is defined as the following: "The degree to which users feel that they understand the AI system or process being explained to them. Compared to Goodness, as defined above, satisfaction is a contextualised, a posteriori judgement of explanations." [32]. Based on previous studies [61] [62] [12], key attributes of explanations can be defined: understandability, feeling of satisfaction, sufficiency of detail, completeness, usefulness, accuracy and trustworthiness. For the scope of this particular research project, the attribute of trust is deemed most significant.

3.4.1 Trust and XAI

Users are prone to perceiving data and claims that systems present to them as valid and accurate because they originate from the computer. But not all users are this easily-convinced. Others might crave a justification of sorts: empirical reasons to conclude that the assertions presented by the computers are true. Different levels of both trust and negative trust (mistrust) exist. The trust users have in automated devices may break down swiftly under certain conditions, for example time pressure or erroneous behaviour by the device [20] [46]. When trust in an automated device is lost, it can be hard to reestablish it.

Trust in the XAI context can be measured in several ways. Either way, the measurement method must take into account the emergence of negative trusting states. The XAI systems should create transparency for the user by enabling them to know if, when and why to trust in that specific system. The same is true for mistrusting the system. Users will invariably have some level of (un)justified trust about the system at hand. The same user might experience different levels of trust for different tasks in the same system [31] [74]. Preferably, the level of trust the user has increases with the experience they have with the system. When this relation, which is based on trust, exists, only then can the user rely on the system with confidence [71].

Previous literature on trust suggests several scales for measuring trust. At a minimum, a trust scale should ask the user these two questions: do you trust the output of the machine, and would you follow its advice? The first question has trust as its target point and the second question is more about the reliance of the system [1]. Others scales, for example the one developed by Schaefer (2013) refers more to the context of human reliance. Does the system act as a part of the team, and is it friendly [75]? Montague (2010) validated a scale for trust in medical diagnostic instruments. The items in the scale referred to the specific medical context (trust in the health care provider and the level of affect towards the provider), but the items also referred to more general items such as reliability, correctness and precision [59]. Results of the studies using scales as the ones described above suggest that trust in automation scales can be reliable. The scales that have been subject to validity analysis, high Cronbach Alpha results have been obtained [34].

3.5 Ecological Interface Design

Cognitive Systems Engineering (CSE) facilitates a framework for analysing, designing and evaluating complex systems [68]. Ecological Interface Design (EID) is an extension of this framework, which targets the usage of interface technologies that are mature. These technologies are used for the development of effective decision-making and problem-solving support. The transformation of the type of behaviour of workers so they can do their job, is a fundamental goal for EID. The understanding of abstract, complex and semantic structures of the domains that people work in is crucial to achieve this goal. Then, an accurate graphical display can be designed to make the work domain both concrete and meaningful. For example, items on the display are easy to locate (concrete) and are easy to understand and make sense of (meaningful). EID was originally introduced by Rasmussen and Vicente in 1989 [69]. Since then, many developments and breakthroughs regarding the theory happened. An overview of the earlier mentioned CSE/EID approach can be found in figure 8. According to this approach, human computer interaction can be classified as a complex and closed-loop dynamical system [7]. An approach such as this one can be called triadic and semiotic: ”The interface is the medium (a “virtual” ecology) that stands between the work domain (situations) and the human (awareness)”. [6]. The foundations of EID (direct perception, direct manipulation and visual momentum) guide the development of interfaces and cut down on the number of design cycles the interface has to run through [6].

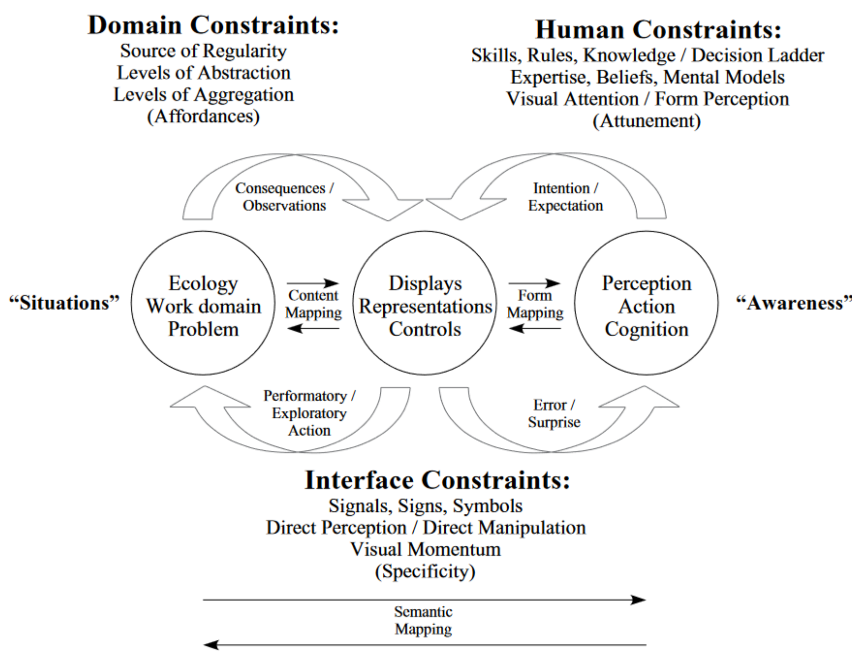

Figure 8: Overview of the CSE/EID approach [6]

The Ecological Interface Design framework attempts to increase and expand upon the perks of Direct Manipulation Interfaces (DMI) in convoluted work domains. Theories of DMI exist and have been researched thoroughly [64]. However, these theories were not designed with the human in mind: the human plays an insignificant role. Yet, humans do play a significant role in the human-machine interaction and it is thus of great importance to involve humans in these kinds of theories. Otherwise, the challenges proposed by the convoluted work domains cannot be addressed properly [69]. To design a more accurate framework for human-machine systems, first the challenges associated with these systems should be identified and clarified [85].

One way to do this is to divide the events in the human-machine system intro groups based on their degree of novelty. This is done from the perspective of both operators and designers. Three groups can be disclosed [85]:

1. Familiar events. These are events that are experienced by the operators frequently. Operators have developed the necessary skills for dealing with these kinds of events because of their accumulated experience and training.

2. Unfamiliar but anticipated events. These kinds of events transpire infrequently. As a result, operators cannot rely on their experience to deal with the events. Since the events have been anticipated on by the designers of the domain, the equipment and instruments (such as DSS and certain procedures) to deal with them already exist. This can aid the operators when coping with unfamiliar events.

3. Unfamiliar and unanticipated events. These events occur rarely and are thus unfamiliar for the operators. More importantly, the events have not been anticipated on by the designers, which results in a situation where the operators cannot rely on previously developed solutions or aids, so it is completely up to them to come up with solutions.

In an ideal situation, unanticipated events in complex domains do not occur at all. This would ease the stress and burden on the operators completely. However, studies show that this is not a realistic situation; it is not technically feasible [45]. A chance will always exist that crucial events that are a threat to the safety will be overlooked and as a consequence will not be included in the design of the domain [58].

The performance of human operators in routine events is sensitive to errors [65]. On the other hand, performance of human operators in unanticipated events is more limited by mistakes, such as an error of intention. The error frequency can be dwindled by applying traditional ergonomic guidelines. However, errors can be prevented only when taking into account the cognitive factors that have an influence on the behaviour of the operator [92]. As a result, it becomes clear that challenges proposed by unanticipated events cannot simply be solved by developing a clear interface with a good layout and well-labelled controls and displays. However, many research that has been done into this area focuses on these types of issues; ergonomic issues rather than semantic issues. Thus, traditionally developed interfaces fail to provide operators with the appropriate amount of support that is needed to deal with unanticipated events [85].

In EID, five key principles exist: Direct Perception, Law of Proximity Compatibility, Skill-, Rule-, and Knowledge-based Behaviour, Abstraction Hierarchy and Transparency [85]. Direct Perception entails that information is presented in a way that is easy for users to understand. They should immediately understand what the state of the system is, and how its dynamics work, without processing complex interactions or complicated thought processes. This principle can be applied in either visualisations or affordances. With visualisations, the user is immediately made clear what the physical process or system status is. An example of this is the usage of a thermometer to indicate the temperature in weather apps. Affordances relate to controls and displays that are designed in a way that immediately shows what their function is. This could be a slider that is used as a controller for adjusting the volume on a Windows PC.

The Law of Proximity Compatibility has everything to do with the placement of items in an interface. Items that needs to be compared or integration, should be located closely together. This reduces cognitive load. Item should be both located closely together and grouped so that the right items are in close vicinity of one another. An aviation cockpit is an example of Proximity Compatibility. Important parameters such as altitude and airspeed are located closely to each other on their dashboard, and are displayed in a way that grabs the attention.

It is important that interfaces encourage different degrees of cognitive control. This can be done by placing the focus on skill, rules, or knowledge (or a combination of the three). Skill-based behaviour is about actions that happen automatically to the user; these are also defined as routine actions. Rule-based behaviour are actions that are based on rules that the user has learnt, or procedures that they know. Knowledge-based behaviour is about problem-solving and reasoning. This is especially useful in situations that are novel for the user.

When Abstraction Hierarchy is applied, information is organised in such a way that users understand what is happening and why. This is done by adding hierarchy to displays, and context to information. By adding hierarchy, users can better navigate through the interface, from high-level overviews to detailed information. When providing context to this information, user better understand the implications of the information they have just seen.

Lastly, transparency is utilised to provide the user with more insight into the state of the computer, how it works and how it behaves. In the interface, the constraints, opportunities and limitations of the system should be clear to the user. This can be achieved by displaying constraints such as safety limits on a place in the interface that is directly accessible to users, or by showing the relationships between different variables.

4 Methodology

4.1 Research Aim

For this study, it will be investigated whether added features to the IGMA system increase trust and transparency. This is done because attaining more insight into the levels of trust participants have towards the system can increase the odds of them actually using the system. Moreover, it can also provide insight into responsible usage. This can, in the end, increase the odds of usage as well. When trust and responsibility levels are low, odds of a successful implementation are expected to be low, as well. To attain these insights, control room personnel will provide feedback about their expectations and concerns regarding the tool via a think-aloud session. Participants will be shown both regular IGMA and IGMA with added features. They will speak their mind about sketches of the IGMA system. Two versions of the added features exist, making this an A/B/C-study. The added features will either be of textual nature or visual. Research [32] [62] has shown that several types of explanations, within the context of AI, exist. The two most important types are visual and textual explanations. Both of these explanations can help user to feel more acquainted with the system and trust more easily. These two types, in combination with two techniques of Ecological Interface Design [85], will be utilised to gain insights into the preference of users.

4.2 Design of the Evaluation

A qualitative research method, namely A/B/C-testing, was applied to this research. A/B/C-testing comes down to the evaluation of three versions of a system: the "regular" version (A) and versions that contain added or modified features (B/C). This way, it can be tested which version of the system performs better or is preferred in certain contexts. Next to a control version (the regular version with no added features), treatment versions (the versions that have been modified) were developed. In this case, the features that are added to the B and C-version of the system were based on the literature study.

Before the study started, an initial interview was conducted to find out more about which exact feature was best to add to the system. The aim of this part of the research was to obtain more information about the current system, and which features were lacking in this system. Then, these features could be added to the prototype to test the trust and transparency levels with. However, it turned out that the operator had interesting thoughts to share about the way of working with their current systems, but very little thoughts on how to improve it or where flaws were located. To summarise: she just used the software, and was not aware of any improvements or flaws. Thus, the data was not incorporated into the eventual prototype.

4.3 Variables

In this study, both independent and dependent variables are used. A distinction exist between independent and dependent variables. Independent variables are not influenced by other variables in a study; hence the term "independent". These variables usually cause changes in the dependent variable. In this study, three independent variable exist: version A, B and C of IGMA. Version A is the control version, version B has added textual explanations and version C has visual explanations. The modifications to the system are features that were added to IGMA, based on the literature. Also, dependent variables exist in the study. The dependent variables are trust and transparency. The first variable, trust, measures the level of trust the participants have in the system. This was tested through the think-aloud sessions with questions such as "do you trust this feature" or "do you feel like this system makes the right decisions?". Transparency was measured through the think-aloud sessions as well, in a similar way. The participant was asked questions about the level of transparency they perceive in the system. An of such a questions is "do you understand why the system makes this decision?".

4.4 Ecological Interface Design

To make certain sections of the interface stand out more, principles of Ecological Interface Design (EID) were used [85]. Two important principles are utilised to make the interface as efficient and easy to use as possible: the Law of Proximity Compatibility and Abstraction Hierarchy.

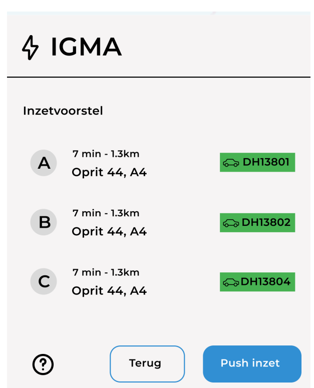

Figure 9: Window with question mark that opens up new interactions

4.4.1 Law of Proximity Compatibility

Firstly, the law of proximity compatibility was utilised to make the interface as efficient as possible. This law entails that information that is to be processed together/simultaneously, should located in close proximity to each other. In this case of IGMA and the visual/textual explanations, the decision was made to place the explanations directly under the piece of interface that already facilitates action and information; the window that allows the user to pick a deployment proposal. As can be seen in figure 10, the two windows that facilitate interaction are located in close proximity. This way, the user does not have to scan through the entire interface to find the information or interaction they need, which increases the efficiency and usability of the software.

Figure 10: Proximity: locate two windows close to each other

4.4.2 Abstraction Hierarchy

Secondly, abstraction hierarchy is utilised to create a difference in the information that is displayed. In this case, the section of the interface that required attention was brought more to the front of the interface, by slightly blurring the rest of the application. The boxes with the pre-conditions (either textual or visual), can be seen by clicking on the question mark. This is done deliberately, because having a great amount of information statically on the screen is not beneficial in the context of the control room: decisions need to be made quickly and the software should be used to interact quickly with. When users click on the question mark, the rest of the interface becomes blurry and the pre-conditions are brought to the attention of the user. In figure 11, the interface with the visual cues can be found. In figure 9 , the question mark is shown that needs to be clicked to open the small screens for the pre-conditions.

Figure 11: Visual explanation with EID technique: (subtle) blurring

4.5 Evaluation Method

The options for evaluating a system in a context like this one, turned out to be limited. Because the situation in the control rooms is ever-changing, it is difficult to create situations that are reproducible. Creating tests is simply not possible, since there often is no right answer to a case. Thus, a think-aloud session was deemed the most useful, because operators can freely share their thoughts. The goal of these sessions was to attain as much information as possible, within certain guidelines, to create a more accurate picture of the patterns of use of the IGMA system and the proposed added features.

A great deal of research has been done on the topic of evaluation. This research has proven to be insightful, and observations can be implemented based on this research. For example, thinking-aloud turned out to be more a more accurate method of identifying issues with systems related to usability than interviews, questionnaires and the analysis of data logs [30]. Furthermore, using other methods of data analysis in combination with thinking-aloud sessions, were found to have a limited added value to the actual analysis [84]. Thinking-aloud sessions can be used to diagnose issues with several IT-related technologies, like the interface of a website [5], which is relevant for this study. Interviews, and more particularly semi-structured interviews, are a better choice when there is a need for user feedback on a more general topic. When testing specific features of a website or other interface, thinking-aloud has proven to be more insightful [22] [19]. Questionnaires are good tools to use when problem areas need to be defined, but these techniques are inadequate when the approach is more open-ended [30]. Also, questionnaires are not an appropriate medium to use when there's a need for input on the design of a system [21]. These techniques only elicit a small percentage of the critical issues previously defined by thinking-aloud [21]. Based on the literature described above, it can be concluded that thinking-aloud sessions are an appropriate technique to detect opinions and problems about certain features.

4.6 Materials

Not a great variety of materials were necessary to conduct this research. FIGMA was used as software to create the sketches and prototypes in. These sketches were shown to the participants in the think-aloud session on a laptop. Google's recording software on their flagship smartphone was used to record the sessions, and later the pre-installed voice recognition software was used to transcribe the data. However, these transcriptions were flawed and had to be checked and improved in their entirety, to make sure the meaning of the words spoken by the participants and the interviewer were correct.

4.6.1 Treatment Version B: Textual Explanation

For the first altered version of the system, a textual explanation was added, as shown in figure 14. The text should clarify more about the suggestions of the IGMA system; it explains the pre-conditions that exist for the suggestions.

Figure 12: Textual Explanation of the Pre-conditions

Figure 13: Textual Explanation of the Pre-conditions (zoomed in)

4.6.2 Treatment Version C: Visual Explanation

The second altered version, the C-version of the IGMA, contained visual explanations, based on the textual explanations of the B-version. This is shown in figure 15. Visual cues were chosen since they can be used to quickly obtain information in a stressful situation, like the situations common in the control rooms [85].

Figure 14: Visual Explanation of the Pre-conditions

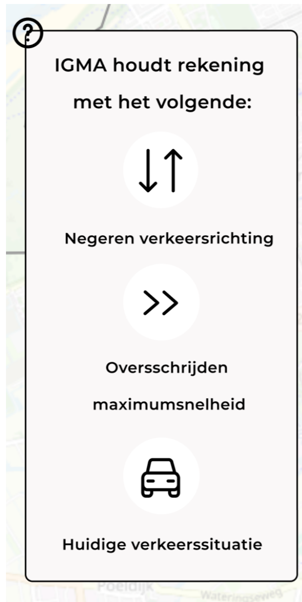

Figure 15: Visual Explanation of the Pre-conditions (zoomed in)

4.7 Participants

A total number of 6 control room operators participated in the study. Of these 6 participants, four were from the control room in Amsterdam (one female participant and three male participants), and two from the control room in the Hague (two female participants). Two of the four Amsterdam operators were regular police operators, and the other two were both police and fire department operators. From the Hague, the two participants were both regular police operators. To be an operator for both the police and the fire department entails that these operators take calls that can be both of police or fire department nature. In practice, it does not affect their duties in a significant manner: the fire department-related calls are of a much smaller frequency than the police-related calls. Participants were in the age group of 30-55 and all had experience of at least three years.

4.8 Procedure

The think-aloud sessions took place offline, in the control rooms (two interviews took place in a separate room and two other interviews took place in the actual control room). After introducing the interviewer and what the research was about, participants were asked to (briefly) introduce themselves and explain what their experience is in the field. Afterwards, they were shown the informed consent en were asked their permission to start the research and record the audio through a smart-phone. One participant gave no permission to record the session, since the interview took place in the control room. The exact location of the participant was in close vicinity of a Lead Operator that often discloses confidential information. Thus, recording the audio might compromise certain situation and was therefore deemed better to not record.

Then, the semi-structured think-aloud session began. These sessions were conducted in the Dutch language. Participants were shown a sequence of 6 screens, and were asked their opinions about the screens. The aim of the questions was to gain more insight into the participant's trust and understanding of the screens shown to them. Since the sessions were semi-structured, lots of possibilities arose to ask more questions to delve deeper into participant's feelings and opinions. The think-aloud session took 35 - 60 minutes. After completing the sessions, participants were thanked and a small coffee break was initiated to chat about the procedure and their work in general. The entire procedure can be found in figure 16.

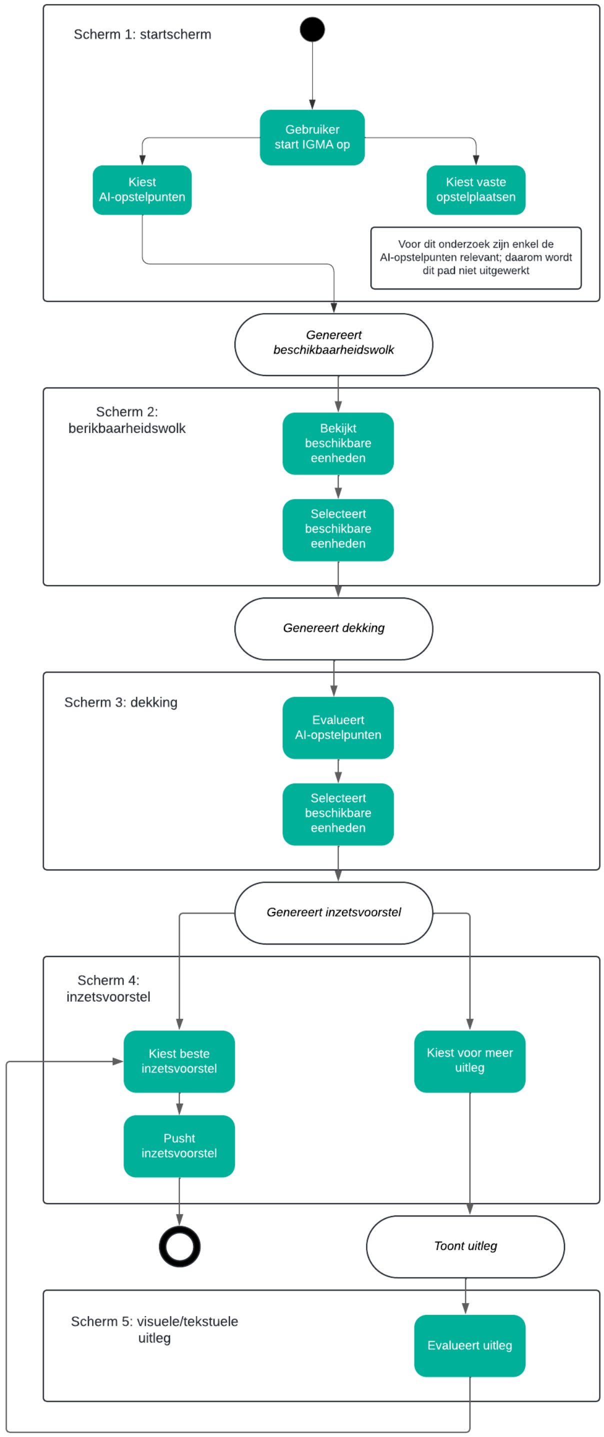

Figure 16: Activity diagram of the IGMA system

4.9 Data collection and analysis

Since the yielded data was all qualitative of nature, statistical analysis was not possible. Therefore, the qualitative data had to be categorised in a way that made it possible to analyse. As it turned out, classifying the statements into themes made analysis possible.

The think-aloud sessions, of which the audio was recorded, were transcribed so that analysis was made possible. Reading the transcripts helped with the familiarisation of the content. After thorough readings and analysing all the statements, 5 themes stood out: time, convenience, integration, training and trust. The statements were categorised into these 5 theme's to back up the overarching implication of the theme.

4.10 Difficulties

Several difficulties were encountered while this research was conducted. First and foremost, finding participants turned out to be challenging. Since control room operators have jobs that require their full attention during their entire workday, it was difficult to arrange a time in their busy schedules to participate in the research without being away from their duties for too long. Moreover, control rooms are places unavailable to the public. While some contact details were obtained through connections within the police force, getting in contact with actual control room operators was demanding. Nevertheless, with a lot of persistence and keeping in contact with employees for a while, a total number of 6 operators were found able and willing to participate in the study.

Creating a test scenario for an interface designed for a context like this one, was a complicated process. Since (practically) no scenario in a control room is the same, creating one for an interface to be tested in is hard. It was found that it is best to use a hypothetical scenario, in which the actual placement of the deployment locations was not the focus of the interface; the added features were. In the questions that were asked in the think-aloud and interview session, less emphasis was placed on the deployment locations and more on the meaning behind them; did their placement make sense, or whether it was clear that they were placed there.

5 Results

5.1 Introduction

In this section, the results of the study will be analysed. In the study, qualitative data was collected through think-aloud sessions. The context of the study did not allow the collection of quantitative data, since it is not reproducible, testable or quantifiable. Since no quantitative data was collected, the analysis will not involve any statistics. Instead, themes will be used to categorise the data. These themes will be discussed and compared in this section. The think-aloud sessions were utilised to collect as much data as possible. The themes will assist in the process of categorising the big quantities of unorganised, non-categorised data.

5.2 Overview of themes

A systematic approach has been applied to the process of discovering themes. This approach can be found in figure 17. The firs step is to transcribe the audio files using the Google transcribe software. Then, he transcripts need to be improved and then categorised. The categorisation happens per screen: the think aloud sessions were centralised around the 5 screens, and the transcripts can be categorised in these 5 screens to make it more clear which comments are about which screens. The second step is to familiarise one with the data. This is done by carefully reading the transcripts and consulting the audio files when things are unclear. The third step is characterised by coding. The step of coding consists of three types of coding, which makes for three separate coding processes: open coding, axial coding and selective coding. Open coding is about breaking down the data into parts that can be identified, named and categorised. Axial coding is about relating the previously defined parts to one another by identifying relationships. This will result in the creation of categories. Lastly, selective coding is about integrating and refining the previously created categories, which can be grouped into themes in the fourth step of the process. The codes are analysed and can be identified to create separate themes. These themes are then reviewed and checked in the rest of the transcripts. Lastly, a clear definition is required for the themes. They are described into great detail so they are clearly delineated. The last step is to map the themes into visualisations that show the relationship between them. These visualisations can be found in section 5.4 of this chapter. After analysing the data in the way that has been described, five overarching themes were discovered: time, integration, convenience, trust and training. Time relates to the time-constrained context of the control rooms, and that a system needs to be designed with this in mind. Convenience is closely related to time, since with this theme, the time-constrained context is important as well. Only in this case, designing a system that is minimalist and easy to use is important. Integration has been selected as a theme, since participants unanimously mentioned this in the sessions. Lastly, the theme of training stood out from the sessions. This entails that with a certain degree of training, the system could be much more useful.

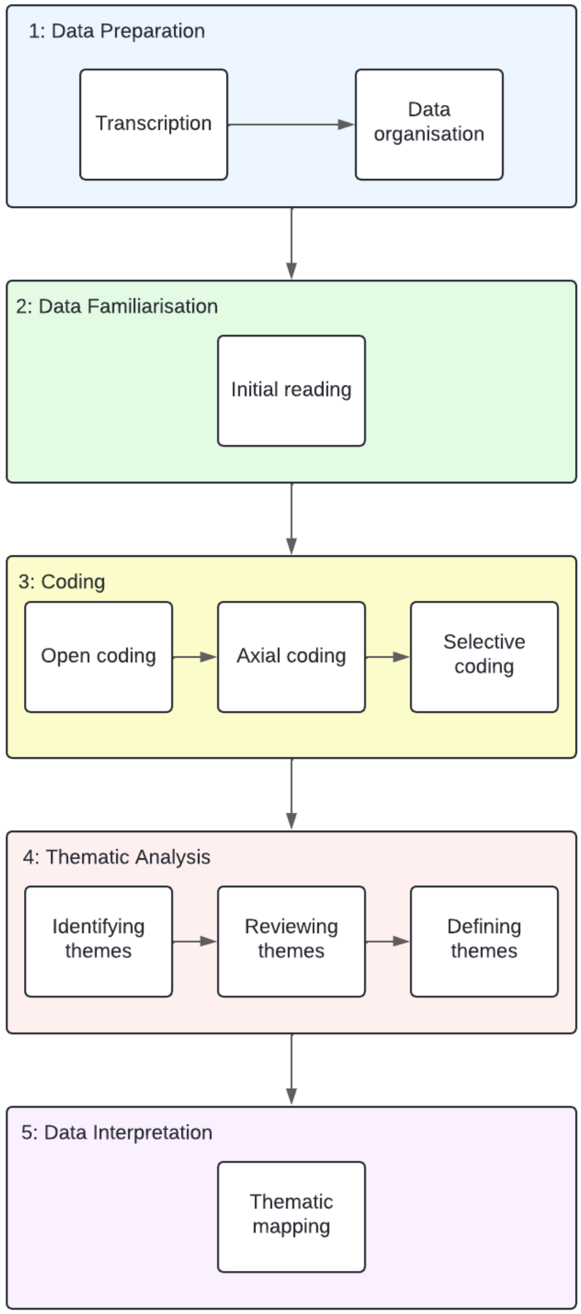

Figure 17: Data analysis process

5.3 Detailed presentation of themes

5.3.1 Theme 1: Time

One of the most important themes, found as a result from the think-aloud sessions, was time. The control rooms are high-paced environments. When a call comes in, a decision needs to be made as fast as possible. It could mean life or death in certain situations. Thus, very little time exists to investigate the call/issue thoroughly. It is important that a system like IGMA takes this into account.

P	Statement
2	Daar heb ik helemaal geen tijd voor, om dat allemaal door te gaan lezen.
3	En dan heb je ook helemaal geen tijd om dit te gaan lezen.
5	Er is geen tijd om je in te lezen in extra informatie.

Table 1: Statements that are related to time

In the think-aloud sessions, screenshots of the IGMA system with the extra explanations were shown. Participants indicated that this would not be useful, since they have very little time as it is. They don't have the time to investigate why IGMA choose certain [opstelplaatsen]. Participants 2, 3 and 5 stated they don't have the time to read this. In the phrases "And then you have no time at all to read this" and the phrase "I don't have time for that at all, to read through all that", more emotions are incorporated. From the sessions, it became clear some participants have developed negative feelings towards the systems they use, since they do not (always) align well with what they need or want. This frustration could be part of the foundation of phrases like these. They seem frustrated with the fact that this system, with its added features, does not take into account the limited time they have in the moment that a call comes in. This would not be beneficial for their eventual adaptation of the system: since they have developed certain feelings towards the system, it might be less likely that they will use it in the end. However, the information itself was appreciated; just not in the heat of the moment when time is so crucial. When this information is incorporated in something like a training, they indicated this would suffice (more on the training in section 5.3.3). Thus, the suggestion to make the system clearer and more transparent is complicated to implement in a high paced context like this one. An explanation is useful, but not when investigating it takes up the precious time of the operator. Important remark: when the participants refer to "this" and "that", when they mention they don't have time to read all that/this, they refer to the visual and textual explanations shown to them in the experiment.

5.3.2 Theme 2: Convenience

Convenience, just as time, is very important in the environment of the control room. When the convenience is as optimised as possible, as little as possible time is lost doing other (unnecessary) tasks. Examples of convenience in the control room are when systems require very little clicks. It is important that when doing tasks to achieve the goal of sending out units, these tasks are as relevant as they can be. Unnecessary clicking and complicated software processes, are not part of this.

P	Statement
1	Je wilt het systeem zo clean mogelijk hebben.
2	Het zou kant-en-klaar op het scherm moeten staan.
3	Als je meer inzicht hebt krijg je meer data op je scherm, dat wil je juist niet.
4	Zo simpel mogelijk houden: hoofd zit zo vol met andere dingen.
5	Dat het bij wijze van spreken met één druk op de knop op het scherm staat.

Table 2: Statements about IGMA's convenience

Participants are fairly content with the degree of convenience the basic form of IGMA offers (without any added features). Actions can be done with a few simple clicks, which is beneficial in this environment. Remarks exist on the integration of the IGMA system with the current system; more on this in section 5.3.4. However, participants are unsure about the usage of IGMA in practice. They wonder whether they have the time to take these extra steps: since IGMA is not necessary in their process of sending out units, they wonder whether they will deem it to be useful in such a way that it outweighs the extra time it takes them. Participants indicate that when the system performs well, they will be probably more enthusiastic about it and see the time-investment as something that is worthwhile. Furthermore, participants indicate the need for less data on the screen. As participant 3 states: with more insights, comes more data on the screen. This is an interesting statement. On one hand, participants indicate that they appreciate more data so they can make decisions that re backed up by more relevant data. And on the other hand, they want to have as less clutter on the screen as possible. It will be a great challenge to find the balance between these two desires. Participant 4 indicates that their head is so full with all the other factors they need to incorporate the moment a call comes in, that they don't have the head space for more data. They want it to be as simple as possible. What participant 2 and 5 indicate, is that it is of great importance that the system is as quick as it possibly can be: with one press of a button, the deployment locations should be displayed on the screen. When minimising clicks, and simplifying the interface as much as possible, these needs can be fulfilled.

5.3.3 Theme 3: Integration

Lastly, the theme of integration emerged from the analysis. Operators use many different pieces of software. The software that they use the most is CityGis. This system incorporates their main workflow: seeing where an incident takes place, checking which units are available and sending the units to the right spots. They also work with other pieces of software, like software that tracks all the calls.

This software is called GMS and is linked to CityGis. Operators are able to click on calls, and see them on the map in CityGis. GMS tracks the calls and prioritises these, as well as software that can be used to track license plates. The different pieces of software all need to be used simultaneously, and take up much space on their three screens. Therefore, it is understandable that IGMA being a separate part of software would not be beneficial in this environment.

P	Statement
1	"...het zou dan eigenlijk wel in CityGis moeten werken."
2	"Het moet ook echt in CityGis draaien, zodat we niet naar nog een scherm hoeven te kijken."
3	"...een minimale eis die wij als centralist stellen: het moet wel in CityGis zitten."
4	"Het moet niet los van CityGis draaien."
5	"...het moet absoluut niet naast CityGis draaien. We hebben al zo veel programma's."
6	"Als het veel ruimte inneemt kan het niet en moet het in CityGis."

Table 3: Statements about the integration of IGMA into other systems

100 percent of participants indicate that they see the odds of usage decline heavily when the IGMA system is not integrated into their current system CityGis. More so, they state that it is an absolute requirement. Participants 2, 3, 4 and 5 explicitly indicate their preference for IGMA to be incorporated into CityGis; participants 1 and 6 also indicate this, but are more polite and hesitant. Nevertheless, it is clear that IGMA must not be separate software. All participants indicated this themselves; no specific question related to this topic was asked. However, the topic of integration is something that neither the researcher nor the team behind IGMA are responsible for. It is an overarching theme, that needs to be taken seriously not just by the team behind IGMA, but more importantly, by the people responsible for the IT strategy for the Dutch Police Force.

5.3.4 Theme 4: Trust

An important factor in this study is trust. When do operators trust a system enough to follow its advice? How much clarification do they need? Operators often have a great deal of experience, and their gut-feeling is accurate most of the time. How can a system be developed that can win the trust of these experienced operators? As it turns out, the participants that took part in this study seem to trust the suggestions of the IGMA system without the need for clarifications. Under the guise of "the computer knows best", participants have no doubt that the system gives them recommendations that make sense and are thus to be trusted.

Several participants had something to say about trust-related topics. For instance, participant 1 and 6 state that the computer takes statistics, calculations and other complex things into consideration, which they could never do (especially in the limited time frame). They indicate this saves them a lot of time. Participant 3 agrees; they state that they don't have to think about the complex context of the calculations, which saves them time overall. Participant 5 claims the computer is all-knowing, and would thus know best. This could be

P	Statement
1	Want hij neemt gewoon statistieken en dergelijke waarschijnlijk mee, en dat soort dingen kunnen wij niet in ons hoofd hebben.
2	...dan ga ik er wel vanuit dat natuurlijk gewoon goed berekende punten zijn
4	...hoef je er niet over na te denken.
5	De computer is een soort alwetend iets.
6	Dan is het heel fijn als je gelijk ziet wanneer wat handig is. Scheelt heel veel zelf nadenken.

Table 4: Statements about the trust participants have in IGMA

because the participant is older than the other participants (50+) and is not as familiar with technologies as the other participants. Participant 2 assumes the AI deployment locations are correct, because they assume that the computer does its calculations well.

On the other hand, participants indicate that when the system is faulty and keeps producing suggestions that do not make any sense, they will lose the trust in the system very quickly. Apparently, a thin line exists between trusting the system and disregarding the system as faulty and unreliable. It seems it is the case that the computer is always right and to be trusted, but when one error happens and the system suggests inaccurate deployment locations, all trust is lost and the entire system is abandoned.

5.3.5 Theme 5: Training

In the study, it was investigated whether participants required more information about the AI deployment locations, to increase the odds of usage. Operators work with complex systems, and since time is such an important factor in their work, these systems need to be as easy to use as possible. Things like training can assist in the process of understanding a system. As it tuned out, participants greatly value a training beforehand, so that the system can be used as quickly and efficiently as possible.

P	Statement
1	Als je weet dat hij dat doet, dan hoeft het er niet in weergegeven te worden.
3	Als je gelijk een training daarin hebt gehad, dan ziet het er logisch uit.
5	Het is ook heel belangrijk dat het systeem goed uitgeleerd wordt.

Table 5: Statements about training

Participants indicate that they appreciate the explanations given by the modified IGMA versions, but also state that once they have attained this knowledge, it is not necessary for them to have insight in this information over and over again. As is shown in table 5, participants state that the information is useful and can help to better understand the system, but also that it is unnecessary to have access to this information after the training. Participant 5 indicates the importance of training overall. They state that it is of great importance that working with a system is actually taught. That time is created in their schedules, so they can earn how to handle the system. They indicate this would be greatly appreciated since the systems are often complex. Participant 1 states that once they have attained the knowledge about how the system works and what their recommendations are based on, they don't have to see this once again since it does not change. If they know, they know. Lastly, participant 3 had doubts about the actual meaning of the AI deployment locations in the screens they were shown. When the meaning was explained afterwards, they indicated that it made sense once they knew. This means that the system overall can make much more sense when the time is taken to properly explain it to employees.

5.4 Relationships between themes

Many of the themes can be related to one another. For instance, the relationship between time and convenience is quite obvious. When convenience is greater, actions take less time. So, an increase in convenience has a decreasing effect on time. This results in a positive influence on the factor of time. This is beneficial in the high-paced context of the control room; when actions take less time, decisions can be made more quickly. The relationships also works the other way around. When there is less time, convenience must be as high as possible. If not, decision can not be made quickly which obstructs the processes in the control room. To summarise: when time decreases, convenience should increase. And when convenience increases, time should decrease.

Convenience and integration are closely related, as well. The integration of the IGMA system can greatly decrease or increase the convenience of the IGMA system. When the systems are not integrated (properly), looking of the specific software, presenting it on the screen and then having to walk through the clicking process, is not convenient at all. The systems should be integrated properly, preferably in the current system of CityGis. This way, presenting the user with the AI deployment locations is very convenient and thus takes very little time. The relationship between time, convenience and integration is prevalent here as well. When software is integrated properly, convenience increases, which decreases time.

	Convenience	**Time**	**Integration**	**Trust**	**Training**
Convenience		+	+	-	-
Time	+		+	-	+
Integration	±	±		-	-
Trust	+	-	±		+
Training	±	-	-	±	

Table 6: Relationships between themes

Other relationships exist as well. In table 6, the relationships between the themes are visualised. The meaning of the symbols is as follows: +” for a positive influence, “±” for a moderate influence, and “-” for no direct influence For instance, the relationship between trust and training is interesting to mention as well. Participants mentioned the need for a good and thorough training, so they know where the system’s decisions come from. This enables them to trust the decisions more, since they better understand where they come from.

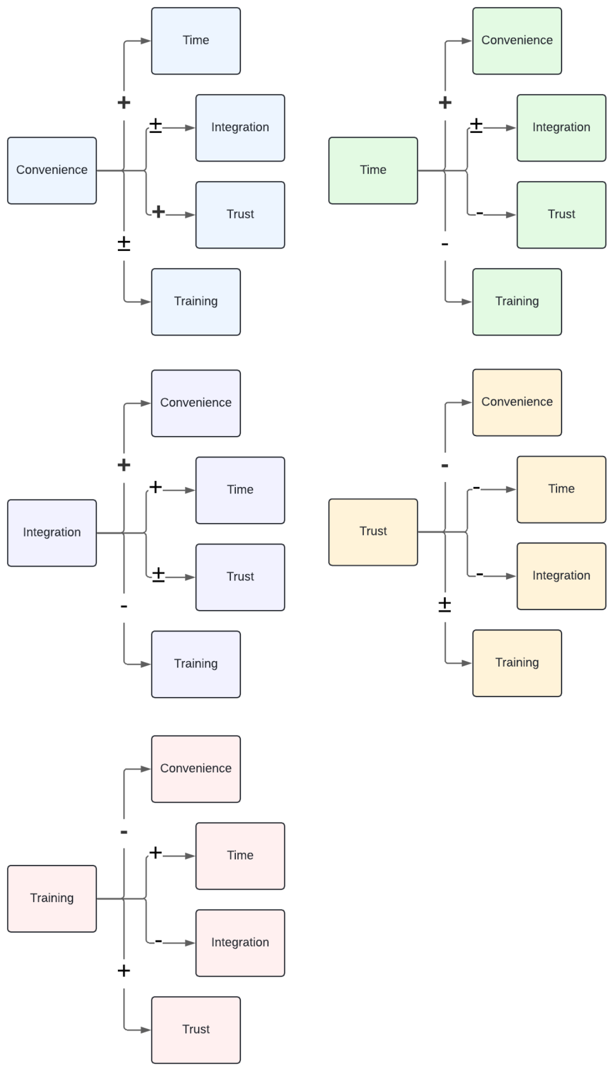

Figure 18: Relationship diagram between the themes

5.5 Unexpected findings

The overwhelming need for the IGMA system to be integrated into current systems, was unexpected. Participants indicated that the odds of them using the system were drastically lowered when it would run as a separate system. This makes sense, since they simply don't have the time to use more systems at the same time.

Furthermore, participants indicated that a high level of cooperation exist between them and the police officers in the cars. Often, officers on the street have more knowledge about certain roads, alleys or other forms of situational awareness. Thus, the operators trust them with finding the right spot to go to. They simply give them the instructions to go to the spot they prefer. Also, a high degree of communication exists between the operators and officers. They talk about the situation, and change their plans according to the feedback they give each other. This is important to incorporate in IGMA, since the AI deployment locations do not allow for much communication: they are static points.

Lastly, the great level of trust the participants had in the system was unexpected, too. It was expected, because of the nature of Police employees, that they would be highly sceptical of a new system and would be more likely to trust their own feelings and experience instead of a system they are not familiar with. This is something that needs to be taken seriously, since it must not be the case that they let IGMA decide without consulting their own thoughts and feelings, first. This topic will be further explored in section 6.1.1.

5.6 Summary of Findings

To summarise, a great amount of data was collected though the think-aloud sessions. This data was then analysed and categorised into 5 themes: convenience, time, integration, trust and training. The themes can be related to each other in different ways. Some influence each other, whiles others don't. The 5 themes are important factors to take into account when developing a Decision Support System like IGMA. The most important takeaways are that the IGMA system must be integrated in the current software, must be easy to use, must be time-efficient to use, is easily trusted since the computer is seen as an all-knowing entity, and that the system must be taught well through training to fully optimise its potential.

The initial goal of the experiment was to investigate whether a visual or textual explanation would increase the trust of the participants. As it turned out in the think-aloud sessions, this question was not relevant whatsoever. Participants indicated they they did not have the time to read or view these extra explanations, and the difference between them was not important to them since they were both deemed as irrelevant. The experiment did yield interesting results that could spark discussion, but as with many studies, the results turned out different than was initially anticipated.

6 Discussion

6.1 Findings

6.1.1 Trust: revisited

The most important factor in this study, the essence of the study, is trust. It was to be investigated whether textual or visual explanations can increase the trust participants have in the IGMA system. As it turns out, the trust factor had a different impact on the participants than initially anticipated. It was expected the participants would be sceptical of the suggestions, and would rely on their gut-feeling instead of the systems. The opposite turned was observed: participants indicate that they trust the suggestions of the system, since it is a computer and can think more advanced things and do this quicker than they can on their own. On the one hand, this is a positive observation. When participants trust the system's output, they are likely to use it and take advantage of it. But on the other hand, it is a negative observation. Participants indicate they trust the output of the system, without the system requiring much clarification for the predictions it makes. Participants will experience the recommendations of the system as true and useful, without being critical towards the actual decisions. This could create potentially dangerous situations, since they put too much trust in the system. For example, when the system recommends deployment locations A and B, and the operator, without verifying whether it is actually a good deployment location, sends their units to locations A and B. But in reality, A and B are not correct locations at all. And if the operator had used their experience and gut-feeling, they would have known this. But since they fully trust the system, they stop thinking critically about its decisions. This attitude, in this context, could have impact as far-reaching as, say, the lives of many citizens.

So, what is the actual goal of IGMA then? One could argue the goal of a DSS like IGMA is to fully take control of the decision-making process so that errors are less easily made and the quality of the decisions in general will remain high. However, this is not the goal of IGMA at all. The goal of IGMA is to support operators with suggestions of deployment locations, and eventually deployment suggestions they might not have thought about themselves. Also, the goal is to assist operators that are less experienced and have thus developed less of a gut-feeling, with making the right estimations. IGMA's goal is never to replace an operator and take charge of the decision making process. There is too much at stake here, and people have very valuable experience and feelings that definitely need to be taken into consideration.

6.1.2 Crucial Context

Another greatly important factor in this study turned out to be time. In the context of the control room, time is such a huge factor that it influences everything: every process, every thought, every decision. And in this context, a few seconds could mean the difference between life and death. Since the stakes are so high, it is of great importance that the factor of time is taken seriously and is respected. People's lives could be at stake.

Time, in combination with convenience, greatly defines the context in which a DSS is to be developed. This is challenging, to say the least. These two factors influence each other, and influence every detail of the software and the context that it is used in. The most important take-away is to keep the software as simple as possible, both in the functionalities and design. Actions should take as less clicks as possible and should me placed on the interface in a minimalist fashion. This way, operators can still see the biggest part of the screen where the important things are happening. Also, the usage of lots of text should be minimised. There is no time to read through big chunks of text. Instead, information should be presented visually so that it can be registered in the blink of an eye.

6.1.3 Peripheral Issues

Lastly, for the two factors of training and integration: these themes are surpassing the domain of IGMA and its development. Even though they are to be taken very seriously, since they greatly influence the (successful) adaptation of the software, they are out of the scope of the IGMA project. To add to this, directorial teams have indicated that they want to phase out CityGis, since CityGis is external software. This is not desirable, since the usage of external software makes it challenging to keep the software updated and to keep it innovative. Since CityGis will be phased out, IGMA can not function in this piece of software. Confidently, IGMA can run in other software instead of being deployed as a separate piece. But this is out of the control of the IGMA team. Other teams in the police force decide how to deploy software and how to keep the personnel engaged and learning.

The advice that is given to these teams is to integrate software like IGMA with existing software. Operators have to deal with a great number of different pieces of software already, and clearly do not have the time or the head space for other, separate pieces of software. For the training: participants indicate the need to be trained with new software before its deployment. They do not have the time to learn all about the new software on the go: their work is too high-paced for this and they can not take a break from their calls to learn more about the software. The integration of important details about how the software works, for instance what the recommendations of the IGMA system are based on, are best communicated to the users through training that happens in advance of the deployment.

6.2 Research Questions

Since the eventual outcome of the study turned out to be different than anticipated, the research questions became challenging to answer. In the context of the results that were found, the questions are answered as accurate as possible. But they were written with a different kind of study in mind, so this influences the eventual answers to the questions.

6.2.1 Main Research Question

The main research question for this study was the following: **How can trust and transparency within a criminal escape route visualisation and suggestion system for law enforcement authorities be enhanced to optimise decision-making and user acceptance?** As it turns out, the suggested visual and textual explanations were not sufficient to increase trust and transparency in a suggestion system in the context of law enforcement. Instead, results show that participants have high levels of trust in the system before they start using it: this was unexpected. There seems to be little need for extra explanations, since the computer is seen as a mighty entity that knows everything. However, it is not a solution to just keep the systems as-is and to stop paying attention to the trust and transparency aspects of the software. Since participants indicate that they trust the system easily, and would thus easily deem its recommendations to be true, a dangerous scenario emerges. This is a scenario in which too much trust is place in the system, which could lead to big problems. The goal of IGMA is to assist operators in their work, by coming up with AI deployment locations and, based on these locations, deployment suggestions. The operator can then compare these suggestions to their own ideas and pick the best solution, or combine the two solutions to create an optimal deployment suggestion. However, when participants trust the system so easily, they disregard their own knowledge and trust the system completely with the decisions. This could be potentially dangerous since the system can always make errors. It is of great importance that the users will remain critical towards the system, and truthfully understand where its recommendations come from. This is why more attention needs to be devoted to finding ways that can increase the transparency of the system, without increasing the time and effort it takes to delve into this information.

6.2.2 Research Question 1

The first sub-question of this research was as follows: **How does the level of transparency in the decision-making process impact the perceived trustworthiness of the system?**. The expected results were that when one increases the transparency, it will increase the perceived trust. As shown in the literature study, this has been proven to be true in other contexts. However, this study showed that that knowledge is not applicable to the niche context of the control room. In this context, where time and efficiency are the most crucial factors, the participants indicate that they trust the system without the need for explanation. They do not have the time to read the explanations. For them, the computer is seen as an all-knowing entity that is to be trusted since it is much smarter than the participants. So, to answer the research question: in this case, the level of transparency does not seem to influence the perceived trustworthiness of the system.

6.2.3 Research Question 2

The second sub-question of the study was: **What are the specific information and feedback mechanisms that enhance user understanding of the system's recommendations?** Several techniques exist for creating an interface that is clear and transparent. Information and feedback mechanics are an important factor in this process. For this study, visual and textual explanations were used as an information mechanism to create more transparency around the decisions the DSS makes. Results show that both mechanisms take too much time to read and process, and are thus regarded as irrelevant since they will not be consulted by the operators. For the textual explanation, this is perfectly logical. Operators do not have time to read through big chunks of text and want information to be displayed quickly and in a simple manner on their screens. The visual explanations were thus expected to be the better option, since they can be scanned through fairly easily. However, this technique was also regarded as too time-consuming. An important take-away here is that the operators have extremely little time, and that the feedback and information mechanics thus have to be as minimalist and simple as possible. Suggestions for mechanisms could be to visualise the information in such a way that it is completely ingrained within the software and thus not require any extra clicking or information. The visualisation tool for traffic jams in Google Maps is a good example of this.

6.2.4 Research Question 3

The third and last sub-question of this research was the following: **What factors influence trust in a criminal escape route visualisation and suggestion system among law enforcement authorities?** The outcome of the study was different than anticipated. However, observations were made during the study that can help answer this question. Firstly, it was found that participants easily trust the system, with the argument of it being a computer and knowing more than they do. This was not anticipated, but interesting nonetheless. Secondly, participants indicate that they would trust the system as long as its recommendations are accurate and relevant. From these statements, it can be concluded that a system must be as fault-free as possible, since the trust of participants is easily damaged. Important to note here is that many participants indicated they had to deal with lots of new software over the past period. This could spark a kind of tech-fatigue in them which makes it harder for systems to be trusted by them. They have to deal with so many systems, that are not equally useful, that they can grow tired of the systems easily.

6.3 Limitations

Several imitations exist in this study. Important to note is that these limitations arose outside the control of the researcher. The first imitation is about the participants: the study only had 6 participants. This makes it difficult, impossible even, to generalise the results on a larger scale. Since the researcher was not operative in an actual control room, it was difficult to get in touch with possible participants. To add to this, the police force is an operation that is hesitant to open up to outsiders. This is understandable, since many of their processes are confidential. Because of this, one simply does not walk into a control room to do some experiments. This takes careful communication and planning, so that is the reason only six participants were found ablate to cooperate in the experiment.

To add to this, the participants worked at two control rooms, namely the Hague and Amsterdam. These control rooms are located within the Randstad: the busiest part of the Netherlands. Control room operators from this part of the country have a vastly different way of working and mindset than in the less crowded parts of the Netherlands. This potentially influence the results. Since the participants have a different mindset, for example that they are more open to new pieces of software, it could paint the false picture that all of the control rooms in the Netherlands are this open-minded and positive towards change. This confirms that the results of this study can unfortunately not be applied to the entire country of the Netherlands.

The most important limitation of this study is the nature of the experiment. For this research project, it was not feasible to actually develop a study that could be replicated, with quantifiable results and quantitative data. Creating a case in which answers are right or wrong was impossible: in this context, a right answer just does not exist. The places where the units can be deployed have such a wide range, and while better or worse places do exist, the perfect place does not. Is has everything to do with communication with the units, changing the plan as you go and trying to create the right fit for the situation. And in situations like these, a perfect answers does not exist. To add to this, time is a hugely important factor that makes the creation of a case even more difficult. In the control rooms, the element of time is so crucial that it influences decision making and the way of working in general. In an experiment setting, this crucial factor cannot be replicated. Participants would never feel the same levels of stress and pressure when taking part in an experiment that is created in a setting in which the element of time does not exist.

6.4 Future research

There are many interesting directions that could be taken with this study as the foundation. Firstly, it would be interesting to find out more about the way information could be portrayed in a way that is seamless and does not take up much time. From the research, it could be concluded that both the visual and textual explanations took too much time to read and would therefore not be relevant in the time-crucial context of the control room. Instead, visualisations of these explanations might have a better effect. For example, it could be shown on the roads how busy they are, which implies that this has been taken into account when doing the calculations on the best deployment locations. Google Maps is a good example of this; the software shows where the busy roads are and marks them with colours (green, orange and red) to indicate how busy the roads actually are. Moreover, they add icons to show where accidents happen and where traffic jams are located. Such visualisation techniques might be useful in the context of the control room, since they bled with the system very well and do not take much time to take into consideration.

Secondly, a factor that could definitely use more research is the way one does research in a context like this. It would be interesting to find out more about research in this domain in general. How could one do research? And how would a case be developed that can be applied in this context, with multiple participants, in which right and wrong answers exist so it is testable? In this master thesis project the time and resources simply did not allow research on such a scale, but it would be very interesting to find out more about doing research in domains like these

Thirdly, it would have been interesting to see the differences in control room location. Like mentioned in section 6.3, only the control rooms of the Hague and Amsterdam were involved. It would be interesting to involve some control rooms located more in the rural parts of the Netherlands, so it can be compared to the results yielded in this study.

6.5 Conclusions

To conclude, the course of the study changed drastically during the process of research. Initially, the goal was to find out which type of explanation would be best in increasing the trust of participants. In the end, however, this turned out to not be relevant for participants. They trusted the system regardless of the proposed explanations. Moreover, they found these explanations to be too time-consuming. Thus, the outcome turned out quite differently than anticipated. Nevertheless, it yielded interesting results. Among the five themes (time, convenience, trust, integration and training), discoveries were made regarding the participant's points of view that can be hugely beneficial for the development process of IGMA. Some of these discoveries are that participants trust the system easily, require training beforehand that they can do without it interrupting their work, need the software to be integrated into their current software so they don't get a cognitive overload and have very little time: thus, the system needs to be as simple and efficient as possible.

For the visualisation, other options could be explored in the future. For instance, a slider that indicates how much time has passed. Or a way in which information can be displayed dynamically, so it does not take up much space on the screen at all times. That the type of explanation in this study, the visual and textual explanations, did not lead to the expected results, does not mean every explanation is inefficient and takes too much time. If a way is found to simply and effectively weave the explanations into the software, this could be proven to have great benefits for the operators and thus for the entire control room.

References

[1] Barbara D Adams et al. "Trust in automated systems". In: *Ministry of National Defence* (2003).

[2] Carl Anderson. "Values-based management". In: *Academy of Management Perspectives* 11.4 (1997), pp. 25–46.

[3] Smaranda Belciug et al. "A brief history of intelligent decision support systems". In: *Intelligent Decision Support Systems—A Journey to Smarter Healthcare* (2020), pp. 57–70.

[4] Izak Benbasat and Weiquan Wang. "Trust in and adoption of online recommendation agents". In: *Journal of the association for information systems* 6.3 (2005), p. 4.

[5] Raquel Benbunan-Fich. "Using protocol analysis to evaluate the usability of a commercial web site". In: *Information & management* 39.2 (2001), pp. 151–163.

[6] Kevin B Bennett and John Flach. "Ecological interface design: Thirty-plus years of refinement, progress, and potential". In: *Human factors* 61.4 (2019), pp. 513–525.

[7] Kevin B Bennett and John M Flach. *Display and interface design: Subtle science, exact art*. CRC Press, 2011.

[8] Miroslav Bernkopf. "EFFECTS OF STRESS AND MOTIVATION ON HUMAN RELIANCE ON AUTOMATED DECISION-SUPPORT SYSTEMS". PhD thesis. Monterey, CA; Naval Postgraduate School, 2021.

[9] Anol Bhattacherjee. "Individual trust in online firms: Scale development and initial test". In: *Journal of management information systems* 19.1 (2002), pp. 211–241.

[10] Ann M Bisantz and Younho Seong. "Assessment of operator trust in and utilization of automated decision-aids under different framing conditions". In: *International Journal of Industrial Ergonomics* 28.2 (2001), pp. 85–97.

[11] Robert H Bonczek, Clyde W Holsapple, and Andrew B Whinston. *Foundations of decision support systems*. Academic Press, 2014.

[12] Béatrice Cahour and Jean-François Forzy. "Does projection into use improve trust and exploration? An example with a cruise control system". In: *Safety science* 47.9 (2009), pp. 1260–1270.

[13] Rich Caruana et al. "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission". In: *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 2015, pp. 1721–1730.

[14] Henriette Cramer et al. "The effects of transparency on trust in and acceptance of a content-based art recommender". In: *User Modeling and User-adapted interaction* 18 (2008), pp. 455–496.

[15] Ashley Deeks. “The judicial demand for explainable artificial intelligence”. In: *Columbia Law Review* 119.7 (2019), pp. 1829–1850.

[16] Stephan Diederich et al. “Understanding the impact that response failure has on how users perceive anthropomorphic conversational service agents: Insights from an online experiment”. In: *AIS Transactions on Human-Computer Interaction* 13.1 (2021), pp. 82–103.

[17] Stephen R Dixon, Christopher D Wickens, and Jason S McCarley. “On the independence of compliance and reliance: Are automation false alarms worse than misses?” In: *Human factors* 49.4 (2007), pp. 564–572.

[18] Patricia M Doney and Joseph P Cannon. “An examination of the nature of trust in buyer–seller relationships”. In: *Journal of marketing* 61.2 (1997), pp. 35–51.

[19] Afke Donker and Panos Markopoulos. “A comparison of think-aloud, questionnaires and interviews for testing usability with children”. In: *People and Computers XVI-Memorable Yet Invisible: Proceedings of HCI 2002*. Springer. 2002, pp. 305–316.

[20] Mary T Dzindolet et al. “The role of trust in automation reliance”. In: *International journal of human-computer studies* 58.6 (2003), pp. 697–718.

[21] Maria R Ebling and Bonnie E John. “On the contributions of different empirical data in usability testing”. In: *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*. 2000, pp. 289–296.

[22] Ellie Fossey et al. “Understanding and evaluating qualitative research”. In: *Australian & New Zealand journal of psychiatry* 36.6 (2002), pp. 717–732.

[23] David Gefen, Elena Karahanna, and Detmar W Straub. “Trust and TAM in online shopping: An integrated model”. In: *MIS quarterly* (2003), pp. 51–90.

[24] M Sinan Gönül, Dilek Önkal, and Michael Lawrence. “The effects of structural characteristics of explanations on use of a DSS”. In: *Decision support systems* 42.3 (2006), pp. 1481–1493.

[25] Bryce Goodman and Seth Flaxman. “European Union regulations on algorithmic decision-making and a “right to explanation””. In: *AI magazine* 38.3 (2017), pp. 50–57.

[26] Mark L Graber and Ashlei Mathew. “Performance of a web-based clinical diagnosis support system for internists”. In: *Journal of general internal medicine* 23 (2008), pp. 37–40.

[27] Manish Gupta, Bala Chandra, and MP Gupta. “A framework of intelligent decision support system for Indian police”. In: *Journal of Enterprise Information Management* 27.5 (2014), pp. 512–540.

[28] Peter A Hancock et al. “A meta-analysis of factors affecting trust in human-robot interaction”. In: *Human factors* 53.5 (2011), pp. 517–527.

[29] Steven R Haynes, Mark A Cohen, and Frank E Ritter. “Designs for explaining intelligent agents”. In: *International Journal of Human-Computer Studies* 67.1 (2009), pp. 90–110.

[30] Ron D HENDERSON et al. “A comparison of the four prominent user-based methods for evaluating the usability of computer software”. In: *Ergonomics* 38.10 (1995), pp. 2030–2044.

[31] Robert R Hoffman et al. “Trust in automation”. In: *IEEE Intelligent Systems* 28.1 (2013), pp. 84–88.

[32] Herwig CH Hofmann and Isabella Lorenzoni. “Future Challenges for Automation in Competition Law Enforcement”. In: *Stan. Computational Antitrust* 3 (2023), p. 36.

[33] Daniel Holliday, Stephanie Wilson, and Simone Stumpf. “User trust in intelligent systems: A journey over time”. In: *Proceedings of the 21st international conference on intelligent user interfaces*. 2016, pp. 164–168.

[34] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. “Foundations for an empirically determined scale of trust in automated systems”. In: *International journal of cognitive ergonomics* 4.1 (2000), pp. 53–71.

[35] Grover S Kearns et al. “Instructor’s manual with test bank to accompany decision support systems: a knowledge-based approach.” In: (1996).

[36] Buomsoo Kim, Jinsoo Park, and Jihae Suh. “Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information”. In: *Decision Support Systems* 134 (2020), p. 113302.

[37] René F Kizilcec. “How much information? Effects of transparency on trust in an algorithmic interface”. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2016, pp. 2390–2395.

[38] Sherrie YX Komiak and Izak Benbasat. “The effects of personalization and familiarity on trust and adoption of recommendation agents”. In: *MIS quarterly* (2006), pp. 941–960.

[39] Chih-Hao Ku and Gondy Leroy. “A decision support system: Automated crime report analysis and classification for e-government”. In: *Government Information Quarterly* 31.4 (2014), pp. 534–544.

[40] Todd Kulesza et al. “Too much, too little, or just right? Ways explanations impact end users’ mental models”. In: *2013 IEEE Symposium on visual languages and human centric computing*. IEEE. 2013, pp. 3–10.

[41] Fujun Lai, Dahui Li, and Chang-Tseh Hsieh. “Fighting identity theft: The coping perspective”. In: *Decision Support Systems* 52.2 (2012), pp. 353–363.

[42] Nancy K Lankton, D Harrison McKnight, and John Tripp. “Technology, humanness, and trust: Rethinking trust in technology”. In: *Journal of the Association for Information Systems* 16.10 (2015), p. 1.

[43] John D Lee and Katrina A See. “Trust in automation: Designing for appropriate reliance”. In: *Human factors* 46.1 (2004), pp. 50–80.

[44] Roy J Lewicki and Barbara Benedict Bunker. *Trust in relationships: A model of development and decline.* Jossey-Bass/Wiley, 1995.

[45] M Lind. “Representations and abstractions for interface design using multilevel flow modeling”. In: *Human-Computer Interaction and Complex Systems* (1991), pp. 223–243.

[46] Poornima Madhavan and Douglas A Wiegmann. “Effects of information source, pedigree, and reliability on operator interaction with decision support systems”. In: *Human factors* 49.5 (2007), pp. 773–785.

[47] Poornima Madhavan, Douglas A Wiegmann, and Frank C Lacson. “Automation failures on tasks easily performed by operators undermine trust in automated aids”. In: *Human factors* 48.2 (2006), pp. 241–256.

[48] Dietrich Manzey, Juliane Reichenbach, and Linda Onnasch. “Human performance consequences of automated decision aids: The impact of degree of automation and system experience”. In: *Journal of Cognitive Engineering and Decision Making* 6.1 (2012), pp. 57–87.

[49] Christian Matt and Constantin Schlusche. “How Transparency Measures Can Attenuate Initial Failures of Intelligent Decision Support Systems”. In: *ECIS 2022 Proceedings* (2022).

[50] Roger C Mayer, James H Davis, and F David Schoorman. “An integrative model of organizational trust”. In: *Academy of management review* 20.3 (1995), pp. 709–734.

[51] Deborah L McGuinness and Paulo Pinheiro Da Silva. “Explaining answers from the semantic web: The inference web approach”. In: *Journal of Web Semantics* 1.4 (2004), pp. 397–413.

[52] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. “Developing and validating trust measures for e-commerce: An integrative typology”. In: *Information systems research* 13.3 (2002), pp. 334–359.

[53] D Harrison McKnight, Larry L Cummings, and Norman L Chervany. “Initial trust formation in new organizational relationships”. In: *Academy of Management review* 23.3 (1998), pp. 473–490.

[54] D Harrison Mcknight et al. “Trust in a specific technology: An investigation of its components and measures”. In: *ACM Transactions on management information systems (TMIS)* 2.2 (2011), pp. 1–25.

[55] Stephanie M Merritt and Daniel R Ilgen. “Not all trust is created equal: Dispositional and history-based trust in human-automation interactions”. In: *Human factors* 50.2 (2008), pp. 194–210.

[56] Christian Meske and Enrico Bunde. "Transparency and trust in human-AI-interaction: The role of model-agnostic explanations in computer vision-based decision support". In: *Artificial Intelligence in HCI: First International Conference, AI-HCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*. Springer. 2020, pp. 54–69.

[57] David M Messick. "Alternative logics for decision making in social settings". In: *Journal of Economic Behavior & Organization* 39.1 (1999), pp. 11–28.

[58] Richard A Miller. "Identification of finite state models of human operators". In: *Systems Research Group. Dept Industrial & Systems Eng, The Ohio State University, Columbus, March* (1979).

[59] Enid Montague. "Validation of a trust in medical technology instrument". In: *Applied ergonomics* 41.6 (2010), pp. 812–821.

[60] K Mosier and L J Skitka. "Human decision makers and automated decision aids: Made for each other?" In: *Automation and human performance: Theory and applications*. Ed. by R Parasuraman and M Mouloua. Hillsdale, NJ: Erlbaum, 1996, pp. 201–220.

[61] Bonnie M Muir. "Trust between humans and machines, and the design of decision aids". In: *International journal of man-machine studies* 27.5-6 (1987), pp. 527–539.

[62] Bonnie M Muir. "Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems". In: *Ergonomics* 37.11 (1994), pp. 1905–1922.

[63] Manisha Natarajan and Matthew Gombolay. "Effects of anthropomorphism and accountability on trust in human robot interaction". In: *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*. 2020, pp. 33–42.

[64] Donald Norman. "User centered system design". In: *New perspectives on human-computer interaction* (1986).

[65] Donald A Norman. "Categorization of action slips." In: *Psychological review* 88.1 (1981), p. 1.

[66] Scott Ososky et al. "Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems". In: *Unmanned systems technology XVI*. Vol. 9084. SPIE. 2014, pp. 112–123.

[67] K Nadia Papamichail and Simon French. "Explaining and justifying the advice of a decision support system: a natural language generation approach". In: *Expert Systems with Applications* 24.1 (2003), pp. 35–48.

[68] Jens Rasmussen, Annelise Mark Pejtersen, and Len P Goodstein. "Cognitive systems engineering". In: (1994).

[69] Jens Rasmussen and Kim J Vicente. “Coping with human errors through system design: implications for ecological interface design”. In: *international Journal of Man-machine Studies* 31.5 (1989), pp. 517–534.

[70] Byron Reeves and Clifford Nass. “The media equation: How people treat computers, television, and new media like real people”. In: *Cambridge, UK* 10.10 (1996).

[71] Victor Riley. “Operator reliance on automation: Theory and data”. In: *Automation and human performance*. CRC Press, 2018, pp. 19–35.

[72] Alan J Rowe and James D Boulgarides. “Decision styles—a perspective”. In: *Leadership & Organization Development Journal* 4.4 (1983), pp. 3–9.

[73] Julian Sanchez, Arthur D Fisk, and Wendy A Rogers. “Reliability and age-related effects on trust and reliance of a decision support aid”. In: *proceedings of the human factors and ergonomics society annual meeting*. Vol. 48. 3. Sage Publications Sage CA: Los Angeles, CA. 2004, pp. 586–589.

[74] NB Sarter, David D Woods, and CE Billings. *Automation surprises*. Vol. 2. New York Wiley, 1997.

[75] Kristin Schaefer. “The perception and measurement of human-robot trust”. In: (2013).

[76] Philipp Schmidt, Felix Biessmann, and Timm Teubner. “Transparency and trust in artificial intelligence systems”. In: *Journal of Decision Systems* 29.4 (2020), pp. 260–278.

[77] A Carlisle Scott et al. “Explanation capabilities of production-based consultation systems”. In: *American Journal of Computational Linguistics* (1977), pp. 1–50.

[78] Younho Seong and Ann M Bisantz. “The impact of cognitive feedback on judgment performance and trust with decision aids”. In: *International Journal of Industrial Ergonomics* 38.7-8 (2008), pp. 608–625.

[79] Matthias Söllner et al. “Understanding the formation of trust in IT artifacts”. In: (2012).

[80] Ralph H Sprague Jr. “A framework for the development of decision support systems”. In: *MIS quarterly* (1980), pp. 1–26.

[81] W Ph Stol and Litska Strikwerda. “Law enforcement in digital society”. In: (2018).

[82] William R Swartout. “Explaining and justifying expert consulting programs”. In: *Computer-assisted medical decision making*. Springer, 1985, pp. 254–271.

[83] Bartosz Swiderski, Jarosław Kurek, and Stanislaw Osowski. “Multistage classification by using logistic regression and neural networks for assessment of financial condition of company”. In: *Decision Support Systems* 52.2 (2012), pp. 539–547.

[84] Lex van Velsen, Thea van der Geest, and Rob Klaassen. "Testing the usability of a personalized system: comparing the use of interviews, questionnaires and thinking-aloud". In: *2007 IEEE International Professional Communication Conference*. IEEE. 2007, pp. 1–8.

[85] Kim J Vicente and Jens Rasmussen. "Ecological interface design: Theoretical foundations". In: *IEEE Transactions on systems, man, and cybernetics* 22.4 (1992), pp. 589–606.

[86] Ewart de Visser and Raja Parasuraman. "Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload". In: *Journal of Cognitive Engineering and Decision Making* 5.2 (2011), pp. 209–231.

[87] William P Wagner. "Trends in expert system development: A longitudinal content analysis of over thirty years of expert system case studies". In: *Expert systems with applications* 76 (2017), pp. 85–96.

[88] Weiquan Wang and Izak Benbasat. "Attributions of trust in decision support technologies: A study of recommendation agents for e-commerce". In: *Journal of Management Information Systems* 24.4 (2008), pp. 249–273.

[89] Weiquan Wang and Izak Benbasat. "Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents". In: *Journal of management information systems* 33.3 (2016), pp. 744–775.

[90] Weiquan Wang et al. "Effects of rational and social appeals of online recommendation agents on cognition-and affect-based trust". In: *Decision Support Systems* 86 (2016), pp. 48–60.

[91] Severin Weiler, Christian Matt, and Thomas Hess. "Immunizing with information–Inoculation messages against conversational agents' response failures". In: *Electronic markets* 32.1 (2022), pp. 239–258.

[92] David D Woods. "Human factors challenges in process control: The case of nuclear power plants". In: *Handbook of human factors/ergonomics* (1987).

[93] Julia L Wright, Jessie YC Chen, and Shan G Lakhmani. "Agent transparency and reliability in human–robot interaction: The influence on user confidence and perceived reliability". In: *IEEE Transactions on Human-Machine Systems* 50.3 (2019), pp. 254–263.

[94] Feiyu Xu et al. "Explainable AI: A brief survey on history, research areas, approaches and challenges". In: *Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8*. Springer. 2019, pp. 563–574.

[95] Jingjun David Xu, Ronald T Cenfetelli, and Karl Aquino. "Do different kinds of trust matter? An examination of the three trusting beliefs on satisfaction and purchase behavior in the buyer–seller context". In: *The Journal of Strategic Information Systems* 25.1 (2016), pp. 15–31.

[96] Jessie Yang, Christopher D Wickens, and Katja Hölttä-Otto. "How users adjust trust in automation: Contrast effect and hindsight bias". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 60. 1. SAGE Publications Sage CA: Los Angeles, CA. 2016, pp. 196–200.