10-9-2023

AI in Government: A Study on Explainability of High-Risk AI-Systems in Law Enforcement & Police Service

Fabian Walke
FernUniversität in Hagen, Germany, fabian.walke@fernuni-hagen.de

Lars Bennek
FernUniversität in Hagen, Germany, lars.bennek@studium.fernuni-hagen.de

Till J. Winkler
FernUniversität in Hagen, Germany, till.winkler@fernuni-hagen.de

AI in Government: A Study on Explainability of High-Risk AI-Systems in Law Enforcement & Police Service

Research Paper

Fabian Walke[1], Lars Bennek[1], Till J. Winkler[1,2]

[1] FernUniversität in Hagen, Chair of Information Management, Hagen, Germany
[2] Copenhagen Business School, Department of Digitalization, Frederiksberg, Denmark
{fabian.walke,till.winkler}@fernuni-hagen.de; lars.bennek@studium.fernuni-hagen.de

Abstract: Law enforcement and police service are, related to the proposed AI Act of the European Commission, part of the high-risk area of artificial intelligence (AI). As such, in the area of digital government and high-risk AI systems exists a particular responsibility for ensuring ethical and social aspects with AI usage. The AI Act also imposes explainability requirements on AI, which could be met by the usage of explainable AI (XAI). The literature has not yet addressed the characteristics of the high-risk area law enforcement and police service in relation to compliance with explainability requirements. We conducted 11 expert interviews and used the grounded theory method to develop a grounded model of the phenomenon AI explainability requirements compliance in the context of law enforcement and police service. We discuss how the model and the results can be useful to authorities, governments, practitioners and researchers alike.

Keywords: *XAI, Artificial Intelligence, AI Act, Requirements, Compliance.*

1 Introduction

Artificial Intelligence (AI), such as ChatGPT and other advanced AI models, have taken the topic of AI and machine learning (ML) to a new level, attracting the interest of both experts and laypeople. The term AI encompasses various research areas, including ML, natural language processing (NLP), speech, vision, and robotics (Mukhamediev *et al.*, 2022), without having a standard definition (Buxmann and Schmidt, 2019). Politicians have also taken notice of this technology, which is shown through the publication of the AI strategies of the European Commission (2018) and the German federal government (2020), and the proposal of the 'Artificial Intelligence Act' (AIA) by the European Commission (EC) (2021). The AIA categorizes AI into four different risk levels: unacceptable risk, high risk, low risk, and minimal risk. Unacceptable-risk systems are prohibited, and high-risk systems must comply with various requirements, including transparency and interpretability. Law enforcement authorities, which include state police authorities, federal police, customs investigation authorities, tax investigation, the federal criminal police office and prosecutors' authorities, are included in the high risk

18th International Conference on Wirtschaftsinformatik,
September 2023, Paderborn, Germany

sector according to AIA Appendix III. The ethics guidelines of the EC (2019) go into more detail about the requirements: The entire decision-making process must be recorded and documented, and an "understandable explanation of the algorithmic decision-making processes" is required, as far as possible. Art. 13 (1) AIA stipulates with regard to transparency that users must be able to interpret and use the results appropriately. The corresponding field of research is called 'Explainable Artificial Intelligence' (XAI), which focuses on ensuring transparency and interpretability of AI systems. The high-risk sector is subject to the explainability requirements of the AIA, making law enforcement and police services (LEPS) subject to XAI and explainability requirements. XAI is of great importance for digital responsibility, as it allows decision-making processes of AI systems to be traced and understood, ultimately leading to greater accountability and transparency. Additionally, XAI has significant social implications, as it can help mitigate potential risks associated with biases, discrimination, and other negative impacts on individuals and society. In terms of ethical and environmental considerations, XAI can play a critical role in ensuring that AI is developed and used in a manner that aligns with ethical principles and sustainability goals.

So far, the literature has not yet examined the characteristics of the high risk area law enforcement and police service in terms of compliance with AI explainability requirements. The literature has only dealt with topics such as general legal requirements on explainability in ML (Bibal *et al.*, 2021), metrics to measure explainability (Sovrano *et al.*, 2022; Sovrano and Vitali, 2023; Sovrano *et al.*, 2021) and technical and ethical dimensions of XAI (McDermid *et al.*, 2021). This leads to the research question (RQ) of this study: "*Which characteristics determine AI explainability requirements compliance in law enforcement and police service?*" The objective of this paper is to determine those characteristics and represent them in a grounded model.

We applied an explorative and qualitative research approach, using the grounded theory method (GTM), by conducting 11 expert interviews. The grounded theory method and exploratory approach were chosen because this specific research area of law enforcement and police service in terms of AI explainability requirements compliance has not been explicitly addressed in the literature so far. As compliance of AI explainability requirements in the high risk area law enforcement and police services is an emerging topic in the field of digital government, we have an abbreviated review of the literature in this study. In the next sections, we outline our grounded theory method (2.), our findings (3.), we discuss the findings, outline the limitations of this study and the implications for research and practice (4.).

2 Methodology

In this research, we applied the grounded theory methodology (GTM) (Glaser and Strauss, 1967) following the paradigm described by Strauss & Corbin (Strauss and Corbin, 1990; Corbin and Strauss, 2015) and the GTM procedure described by Wiesche *et al.* (2017). The context of this study is the European Union (EU) whereby Germany was selected as the focal case with its multiple embedded units of analysis, represented

by various law enforcement authorities, including police. Germany is a particularly critical case in the EU regarding successful digital government (Walke *et al.*, 2023), because it ranks the penultimate place out of 39 countries in terms of digital competitiveness in Europe and North America (ECDC, 2021) and there are still significant discrepancies between the status quo of e-government services and the requirements of the Online Access Act (OZG) in Germany (Hölscher *et al.*, 2021). By choosing Germany and the high risk area of AI, we applied a critical case selection in our study, as part of an information-oriented selection, which describes the counterpart of random selection. A critical case can be defined as having strategic importance in relation to the general problem and achieves information, that permits logical deductions of the type (Flyvbjerg, 2006). Since in Germany the need for a successful digital government is particularly high and law enforcement and police service are particularly critical areas of AI, we expect characteristics for the grounded model, which could be particularly evident.

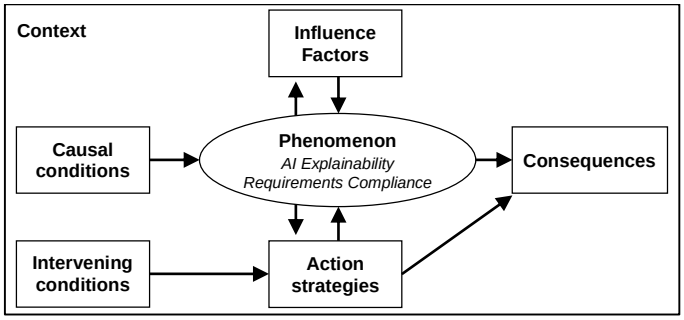

Figure 1. GTM study model, adapted from Corbin & Strauss (2015)

The study pursued an exploratory qualitative research approach by conducting expert interviews and analyzing them qualitatively. The study took place over a period of four months from November 2022 to February 2023. Methodologically, this study was based on nine steps regarding GTM described by Wiesche *et al.* (2017). During the steps of theoretical sampling (1) and role of prior theory (2), and in addition to identifying the research gap, deriving the research questions and data acquisition, we used Strauss and Corbin's GTM paradigm (Strauss and Corbin, 1990) to define a GTM study model and to design our interview guidelines for the expert interviews. Our GTM study model (Figure 1) consists of the categories causal conditions, phenomenon, context, influence factors (divided in general, drivers and barriers), action strategies, intervening conditions and consequences. In our GTM study model the influence factors are referred to the action strategies. To address a wide range of possible action strategies, reference was made within the category of action strategies to the quality dimensions technology, information, human, process and system (Walke and Winkler, 2022; Walke *et al.*, 2023). The experts for the interviews (Table 1) were recruited from the public and private sector. A total of 11 experts were interviewed. The interviews were processed with open and explorative questions related to the given GTM study model category. The prerequisite for selection were expertise in the use, planning or development

of AI systems that are or will be used in the field of law enforcement and police services. The data analysis steps of open coding (3), axial coding (4), selective coding (5), theoretical coding (6), constant comparison (7) and memoing (8) described by Wiesche *et al.* (2017) have been performed using MAXQDA Software and were based on the coding system of the GTM study model (Figure 1). The coding paradigm, the phenomenon (AI explainability requirements compliance) and the relations between the categories have already been predetermined by the GTM study model. The step of the final connection to the coding paradigm (9) was used to develop the final grounded model.

Table 1. Overview of expert interviewees

#	Sector	Area	Position	Work experience	Age
1	Public	Justice	Senior Prosecutor	> 15 years	40-49
2	Private	Public Sector Consultancy	Senior Manager	> 15 years	40-49
3	Private	Public Sector Consultancy	Big Data Scientist	> 3 years	20-29
4	Private	Public Sector Consultancy	Senior Consultant	> 5 years	30-39
5	Public	Research	PhD Candidate	> 5 years	30-39
6	Private	Lawyer Criminal Law	Partner	> 15 years	40-49
7	Private	Public Sector Consultancy	Leading Consultant	> 5 years	20-29
8	Public	Administration of Justice	Head of Unit	> 5 years	40-49
9	Public	Police Authority	IT employee	> 15 years	40-49
10	Public	Administration of Justice	Advisor	> 15 years	40-49
11	Public	Police University	Professor	> 25 years	50-59

3 Findings

We present the findings of this study successively based on the categories of the GTM study model and by separating the findings in three coding steps: axial coding, selective coding and theoretical coding (TC; asterisk* in the following tables means equal to selective coding), especially to provide a logical chain of evidence. The presentation of the coding results is based on the explanations of Williams & Moser (2019). The step of open coding is described textually and provides an excerpt of critical and relevant characteristics that were expressed by the expert interviewees. Additionally, we highlight relevant findings of the general coding process with expressive and direct citations from the expert (abbreviated as E1, E2, E3 etc.).

With the methodological restriction to the field of LEPS, the *context* was determined in relation to the phenomenon. In the course of coding, the special features of LEPS were assigned to the context. E11 emphasizes the higher requirements in law enforcement for the use of AI by describing that errors due to insufficient accuracy could "*completely destroy entire families*". Accordingly, AI systems are also classified as high-risk AI systems in the AIA, which in the opinion of E8 is too sweeping. E6 notes that there are hardly any concrete AI use cases in the area of law enforcement, but that these are "*very much*" discussed. Another special feature in the context of criminal prosecution is the judicial free assessment of evidence according to § 62 of the Code of Criminal

Procedure (StPO), according to which a judge "*can make his own subjective conviction the basis of a guilty verdict on the basis of objective facts*" (E6). E8 adds to judicial independence: "*[...] the judge must not submit to an algorithm. And that is in principle our guiding principle*". The findings regarding context are summarized in Table 2.

Table 2. Context and its coding

Axial Coding	Selective C.	TC
LEPS as a high-risk sector; Low penetration and use of AI in LEPS; Control by judges; Higher requirements in LEPS;	LEPS specifics	*

Within the frame of selective coding, various *causal conditions* were summarized as "Legal AI requirements". AI explainability requirements itself was regularly named as the cause. E10 describes that the AIA specifies in the abstract that an AI decision must be comprehensible and verifiable. E4 emphasizes the importance of XAI metrics to identify misstatements. E6 sees metrics as a decision criterion for AI tool selection: "*Well, at the end of the day, you need some kind of tool*". The findings regarding causal conditions are summarised in Table 3.

Table 3. Causal conditions and their coding

Axial Coding	Selective C.	TC
Objectivity; Decision factor in tool selection; Human limitations; Transparency requirement; Control over AI; AI explainability requirements; Black box; Identify erroneous explanations; Verifiability; Jurisdiction; No additional AIA requirements; European Law; Procedural rights; Judicial free assessment of evidence; Fundamental rights; Duty to state reasons; AIA; Burden of proof;	Legal AI requirements	*

Table 4 lists neutral (general), positive (drivers) and negative (barriers) *influence factors*. A neutral influence was used to code the actors mentioned, such as judges, prosecutors, defence lawyers, defendants, developers and users. Various technological dependencies were also noted, such as metric dependency to use cases, to the AI models or to XAI methods. E9 sees possible positive impacts on AI from technologies such as ChatGPT, as they could increase people's awareness in this regard. The complexity of some AI models is seen as a negative factor. E8 sees the danger, especially with neural networks, that only a "*best possible metric"* is realistic, but he cannot imagine an always correct metric. The availability and maturity of existing metrics was mentioned a total of nine times by five interviewees (E1, 3, 4, 7 and 8). It was noted that the topic was new (E3), that no metrics were known (E4) or that metrics were difficult to find (E8). In addition, bias and potential dependence on technology providers were mentioned. The influence of regulation was assessed differently. In the interviews, both insufficient or no regulation and partly too restrictive regulation were pointed out: "*Yes, and lumping all AI systems of police authorities together is absolutely not helpful*." (E9); "*Let's take a look at these big language models GPT etc. Hardly anything comes from Europe these days. That's just the point. We can regulate everything very well, but then the technology simply moves away and we can't use it as a public sector.*" (E8).

According to E5, clear requirements for a certain level of explainability would also have a driving effect. Regulation can also have a neutral effect if neither a positive nor a negative influence is associated with it. The findings regarding influence factors are summarized in Table 4 and have been condensed during theoretical coding to "Dependencies, "Future perspectives" and "Complexity".

Table 4. Influence factors and their coding

Axial Coding	Selective C.	TC
General Influence Factors (neutral)		
Dependence metric with use case/AI model/XAI-method; Dependence AI model and XAI method; AI model quality criteria; Dependence on technology providers;	Technological dependencies	Dependencies
Abstract requirements; Case law; Basic law;	Regulation	
Injured party; Accused; Professionals; Developer; Experts; User; Defence lawyers; Prosecutor; Judges;	Actors	
Drivers (positive)		
Future perspectives on AI; ChatGPT hype; Clear regulation as a potential driver;	Future perspectives	*
Barriers (negative)		
Lack of legal requirements; Lack of standards; Lack of legal clarity; Restrictive regulation; Data protection;	Insufficient regulation	Complexity
ChatGPT as an example of complexity; Explainability not achievable or meaningful; Paradigm shift/system complexity; Complex models; Limits of measurability; Defining requirements for metrics is complex; Complexity of digital products;	Complexity	
Availability of metrics; Maturity of metrics;	Metrics maturity	
Bias;	Bias	

In the following, the *action strategies* are listed along the five quality dimensions technology, information, human, process and system (Walke *et al.*, 2023; Walke and Winkler, 2022). The action strategies in the *technology* dimension are condensed to "Technological preconditions creation" (Table 5). A compromise between precision and explainability seems necessary. E3 states that in law enforcement, very complex models that deliver very high precision and high recall are not the primary goal. Rather, one tries to find a kind of trade-off between the quality of the model in relation to the current reference system and the simplicity of the explanatory power. This means that one tries to improve the quality minimally, but at the same time to achieve maximum explanatory power and simplicity of the model. Regarding preventing recalculability, E1 mentions: "*So far, we have focused on the question of the technical framework conditions and the recalculability of the AI models in order to ensure that a delivered AI cannot be used later to generate child pornography itself*".

When developing AI models, care must be taken to ensure that they cannot be used to generate prohibited content themselves. An AI that has been trained to recognize specific content could also generate it. Within the *information* dimension and during selective coding we identified "Support intelligibility" as a major action strategy. E2 points out that explainability is there for people and that corresponding metrics must

also be understood by them. In his statement, E1 emphasizes the importance of being able to explain the technical framework conditions in a comprehensible way even to a person who does not have an affinity for technology.

Table 5. Action strategies and their coding

Axial Coding	Selective C.	TC
Technology		
Compromise precision/explainability; Preventing recalculability; Barrier-free; Use transparent models; Do not use ready-made solutions;	Technological preconditions creation	*
Information		
Use of domain language; Understandable explainability; Understandable metrics; Understandable regulation;	Support intelligibility	*
Human		
Acceptance; Error acceptance; Usability; Trust; Highlighting benefits; Publish models used by the state; Create awareness; Create mindset; Motivation; Take away fears; Change;	Promoting acceptance, benefit & trust	AI-focused education
AI Literacy; Data Literacy; Technological understanding; Training; Education;	Building competencies	
Process		
In-house development; Interdisciplinarity; Involvement of relevant stakeholders in the development process; Integrate metrics in the development process; Paradigm shift; Quality assurance; Carry out simulations; Test procedures; Transparency through traceability; Holistic approach;	Development optimization	Normed and optimized XAI development
Monitoring in use; Logging; Retraceability; Versioning; Traceability;	Traceability assurance	
Benchmarks; Best practices (e.g. GDPR); Provide explanatory methods; Recommendations for action; Software repositories; Standardization; Mandatory legal requirements and voluntary standards; Certification; Standardized auditing process;	Requirements through standards	
System		
Exchange with other departments; Promotion of research; Promotion of open source software; Publication of certified systems; Exchange with science; Workshops;	Overarching exchange	Structural and cultural organization
Agility; AI as a central component of an organization; Organizational policy;	Creating AI culture	
Risk management; Auditing unit; Supervision; Federal-state intergovernmental approach; Expert Committee; AI Commission; Taskforce; Certification institute;	Building AI organizations	

One possible action strategy within the *human* dimension is to promote acceptance, benefit and trust. It must be possible to trust the result of a black box (E11). In addition, there must also be a certain acceptance of errors for AI, as it cannot achieve 100% accuracy (E9). Additionally, the action strategy “Building competencies” was identified. E7 assumes that decisions can be made better with a basic understanding of AI.

E8 also argues for a decentralized anchoring of this knowledge and sees the development of digital competences through further training as necessary. We condensed the selective codes during theoretical coding to the action strategy "AI-focused education".

With regard to the *process* dimension and selective coding, we identified the following action strategies: "Development optimization", "Traceability assurance" and "Requirements through standards". According to E8, an interdisciplinary approach that takes both professional and technical aspects into account is necessary to ensure the successful development of XAI. The metrics must already be integrated in the development process (E4) and the same applies to the users of the AI in question (E3). E1 also points out that, in his view, AI can only be used in the justice system if the justice system has a formative influence on it during the development phase, rather than adopting a solution from a manufacturer or scientific team: "*In my opinion, an AI in the justice system can hardly be used in the situation that a manufacturer or a scientific team presents us a solution and says that it can be used in this way; instead, we have to exert a formative influence in the development phase.*" E9 stresses the importance of clearly defined standards to ensure the reliability of AI systems. To this end, he suggests that independent bodies carry out and publish certifications for certain systems in order to achieve standardization. Another important factor is a standardized auditing process carried out by an external institution to check the traceability and reproducibility of the systems. E8 suggests that the definition of standards should focus on both general criteria and specific requirements for certain areas of application. A panel of experts could be used for the general part, while a case-by-case approach would be necessary for specific requirements. The development and updating of the standards could be carried out on the basis of practical experience. We condensed the action strategies of the selective coding during theoretical coding to "Normed and optimized XAI development". Regarding the *system* dimension, we identified during selective coding the action strategies "Overarching exchange", "Creating AI culture" and "Building AI organizations". Aspects such as the promotion of open source software (E5) or research on XAI and metrics (E4) were mentioned here. E1 also emphasized the reason for taking part in the interview, in order to achieve a "*higher degree of formalization also in the area of explainability*" through dialogue and scientific support. With regard to the development of organizational structures for AI systems, a cross-federal state approach is called for (E8) and, for example, a certification institute (E8). We condensed the action strategies of the selective coding during theoretical coding to "Structural and cultural organization". All action strategies are summarized in Table 5.

The action strategies are based on *intervening conditions*, which in turn were coded along the previous quality dimensions technology, human, process and system. Regarding the information dimension, we found no relevant intervening condition. Regarding *technology* and "Technology preconditions", E10 makes clear that AI systems should be comprehensible in order to make their decision-making verifiable. However, it still has to be defined how this traceability is implemented technically. In the framework of the AIA, traceability is taken into account as an abstract requirement, however, there will be no concrete regulations on how traceability is to be technically implemented. Regarding the *human* dimension, "Psychosocial factors" appear as intervening conditions. E7 believes that metrics are particularly necessary in the justice system because

of the principle-based thinking and the high level of responsibility. E8 and E9 state that 100% accurate AI will not exist and that this should be acceptable as humans also make mistakes (E8, E9). In addition, E8 emphasizes that no corresponding metric would exist for humans either, because their decisions themselves are not comprehensible. A *processual* intervening condition is the "Use type of AI". There is the uncritical use as an aid, for example, in the context of file processing (E8) or the decision-supporting use, in which investigations are supported by an AI, but the decision is made by a human (E9). E11 emphasizes: "*[...] the AI itself, no matter how good the metrics are, can never have the sole and final decision.*" E8 states that before the judiciary can enter the high-risk area particularly regulated by the AIA, there is still some groundwork to be done, such as business process automation and metadata extraction. Several expert interviewees stress that the introduction of a metric for XAI raises further questions.

Table 6. Intervening conditions and their coding

Axial Coding	Selective C.	TC
Technology		AI Quality preconditions
Quality of the input data; Robustness; Evaluation of the assessment; Lack of technological basis;	Technological preconditions	
Human		
Rejectionism; Coziness, laziness; Age-related factors; Motivational factors; Avoidance of mistakes; Lack of awareness of the need; People also make mistakes; Human black box;	Psychosocial factors	
Process		
Field of application; Independent decisions; Decision preparation, support; Consequences of decisions; Intended use;	Use-type of AI	
System		
Resources; Cost recovery; Employee availability regarding staff shortage; Lack of time; Expenditure; Intervening regulation;	Resources	

For example, how the system arrives at the underlying metric (E8) or whether there is a need to evaluate the metric: "*[...] there is a category of explainability metrics. Yes, and one says score 80, the other says score 50. Then I will always take the one with the score 80. So the requirement, the driver for us is to keep control of the tool. That is the driver for us as lawyers. And it should also be the driver for the judiciary. And what is an obstacle? Well, of course, the question is then again: don't you also have to assess the metrics? So the assessment of the assessment.*" (E6). Regarding the *system* dimension, E8 wonders, in terms of resources, "*who do you call or where do you get these people?*" At the same time, he sees the opportunities: "*[...] this technology offers great potential to strengthen the state, to strengthen the rule of law, to relieve personnel and to make us really fit.*" (E8). The selective codes have been condensed during theoretical coding to "AI Quality preconditions" (Table 6).

E8 emphasizes that the introduction of an XAI assessment would bring a considerable gain in acceptance, as many reservations on the part of citizens and colleagues could be eliminated as a result. Such a development would also be advantageous from the perspective of the rule of law. E4 sees the possibility of a restriction of AI use if too

high demands are made, including on metrics. These positive, neutral and negative consequences have been condensed during theoretical coding to "Antagonistic impact". The coding of the consequences can be found in Table 7.

Table 7. Consequences and their coding

Axial Coding	Selective Coding	TC
Building trust; Reduction of reservations; Increasing acceptance;	Trust and acceptance	Antagonistic impact
Enabling the appealability of decisions; Compliance with legal requirements;	Legal compliance	
Innovation inhibition;	Innovation inhibition	

Potential metrics were assigned to the category of the *phenomenon* "AI explainability requirements compliance", which is predetermined by the study model (Figure 1). During selective coding we identified "Qualitative metrics", "Quantitative metrics", "Metrics visualization" and "Objectives of metrics". E2 sees a clear need for quantitative metrics, as qualitative metrics would leave the realm of AI and this always meant simplification. E1 can also imagine "*random parallel evaluations"*. The results for the phenomenon can be seen in Table 8.

Table 8. Phenomenon and its coding

Axial Coding	Selective Coding	TC
Charts; Dashboards;	Metrics visualization	AI explainability requirements compliance
Human evaluation, survey; Stress test;	Qualitative metrics	
Anchors; Distance to the tipping point of a decision; Reference corpus / reference model / ground truth; Feature importance; Accuracy; F1 Score; Number of decision core components; Explainable: yes/no; Precision; Recall; Randomization; Meteor; BLEU; Word Error Rate;	Quantitative metrics	
Objectives of metrics;	Objectives of metrics	

Due to the high importance of the measurability of the phenomenon, we additionally compared the literature on XAI metrics (Sovrano *et al.*, 2021, 2022; Sovrano and Vitali, 2023) with the metrics mentioned by the experts. We come to the conclusion that, of the 16 metrics mentioned by the experts, eight could be connected to the metrics mentioned in the literature. The other eight metrics (stress test, anchors, distance to tipping point of a decision, number of decision core components, explainable yes/no, meteor, BLEU and word error rate) have not yet been mentioned in the XAI literature we considered (Sovrano *et al.*, 2021, 2022; Sovrano and Vitali, 2023).

Using the insights gained from the interviews and the subsequent coding, the grounded model (Figure 2) was developed based on the study model (Figure 1), which contains the theoretical coding's.

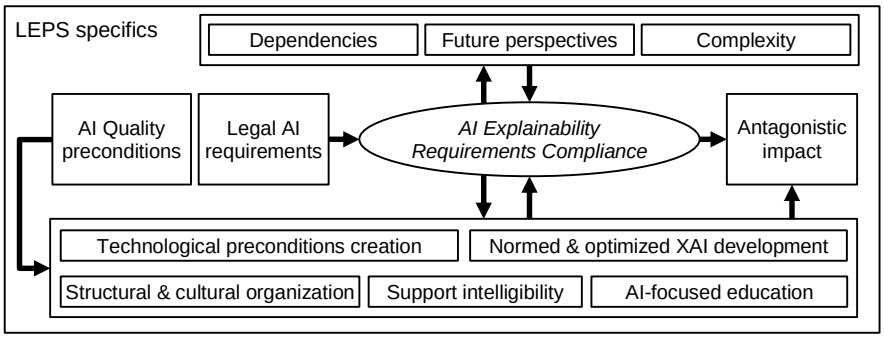

Figure 2. Grounded model of AI explainability requirements compliance

4 Discussion and Conclusion

In this study we followed an exploratory and qualitative research approach regarding the phenomenon—*AI explainability requirements compliance*—using the grounded theory method. The research question was addressed by discovering the characteristics of the phenomenon in the context of LEPS and by representing the characteristics in a grounded model. Dependencies, future perspectives and complexity were identified as key influence factors, and the creation of technological prerequisites, normed and optimized XAI development, structural and cultural organization, supporting intelligibility, and AI-focused education as action strategies. Intervening conditions of action strategies are AI quality preconditions. The cause of the phenomenon are legal AI requirements and the consequence of the phenomenon is an antagonistic impact.

The results of this paper show that the use of AI systems in LEPS is subject to further special requirements, as errors due to insufficient accuracy can lead to serious consequences. It is therefore important that they are designed in such a way that they meet the requirements of the rule of law, the requirements of criminal procedure and, in the future, the requirements of the AIA, while at the same time ensuring a high level of accuracy and reliability. The AI systems used must be transparent and comprehensible in order to ensure that judges and other actors can make their decisions on the basis of objective facts, or can review and challenge them if necessary. The action strategies indicate ways in which compliance with AI explainability requirements can be achieved. Since no metrics in use could be named during the interviews and AI has so far tended to be used in non-critical areas, metrics play a decisive role. The establishment of organizational structures for AI, the creation of certification institutes and, in particular, the provision of standards, are of great importance. Ideally, users must already be involved in the development process. Through co-design, a kind of transparency or comprehensibility can arise even with black-box AI, since the underlying logic can be co-developed and understood. At the same time, AI explainability requirements can be taken into account, including relevant metrics. In this context, it is important that users have the competences to understand the basic features of AI. This can be achieved

by building AI and data literacy competencies through training and education. Awareness of such a need can also be raised through public awareness as a result of technologies such as ChatGPT, as people see what AI can be capable of. Skill enhancement and change management can help foster trust and acceptance in AI systems, providing a foundation for the development of XAI metrics. These metrics can, in turn, help to improve XAI and thus further increase trust and acceptance in these systems. It can furthermore be stated that a use of XAI metrics in the field of law enforcement, at least in the field of interview partners, does not seem to happen yet. AI applications are already in use or in development and explainability also plays a role in these, but their measurement or assessment does not seem to be considered yet.

Within the frame of the free assessment of evidence, a judge must in any case form his or her own opinion about the credibility and reliability of evidence, regardless of whether it comes from AI systems or not. This realization leads to the assumption that metrics can support the judge in this free assessment of evidence. It can be assumed that various degrees of freedom through the combination and free selection of metrics could contribute to this, but ultimately the judge decides to what extent he or she takes these metrics into account. In addition, the question of the explainability of the metric itself was raised, and whether it should not be evaluated itself.

A total of 11 expert interviews were conducted from the judiciary, judicial administration, police, research and private sector, which seemed sufficient to achieve theoretical saturation. However, this limited number of interviewees could be increased in future research. The decision to select LEPS as the critical focal case was based on the recognition that AI systems in this domain are likely to be classified as high-risk AI systems and are therefore subject to particular challenges. However, it could be, that this is precisely why AI systems are still comparatively rarely used in LEPS and thus there may be less expertise and experience in the development and use of AI systems than in other areas. The analysis of another sector could therefore lead to different results with regard to the general aspects. One approach for further research would be to carry out a sector comparison in order to examine the differences and similarities with regard to XAI and their assessment by means of metrics. This could provide insights that go beyond the field of LEPS and may also be applicable to other high-risk AI systems. Specificities in the field of LEPS are highlighted and it is found that challenges exist in the implementation of AI systems and XAI metrics. In order to meet these challenges, the phenomenon was considered holistically by applying a grounded-theory-approach.

This paper examined the specifics of compliance with the explainability requirements of AI systems in the context of law enforcement and police service. In this way, the paper contributes to the discussion about the impact of AI systems on law enforcement and police service and shows that the use of AI systems in this context is associated with special requirements and challenges. The identified characteristics of this high risk area can contribute to the academic discussion, the practical development of AI systems in high risk areas and can help authorities and governments create appropriate conditions for the use of AI.

References

Bibal, A., Lognoul, M., Streel, A. de and Frénay, B. (2021), "Legal requirements on explainability in machine learning", *Artificial Intelligence and Law*, Vol. 29 No. 2, pp. 149–169.

Buxmann, P. and Schmidt, H. (2019), *Künstliche Intelligenz*, Springer Gabler, Berlin, Heidelberg.

Corbin, J.M. and Strauss, A.L. (2015), *Basics of qualitative research: Techniques and procedures for developing grounded theory*, 4th ed., Sage, Los Angeles, Calif.

ECDC (2021), *Digital Riser Report 2021*, ESCP Business School, available at: https://digital-competitiveness.eu/wp-content/uploads/Digital_Riser_Report-2021.pdf.

European Commission (2018), *Künstliche Intelligenz für Europa*, available at: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:52018DC0237 (accessed 28 October 2023).

European Commission (2019), *Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen Schaffung von Vertrauen in eine auf den Menschen ausgerichtete künstliche Intelligenz.*, available at: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:52019DC0168 (accessed 29 October 2022).

European Commission (2021), *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts.*, available at: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:52019DC0168 (accessed 29 October 2022).

Flyvbjerg, B. (2006), "Five Misunderstandings About Case-Study Research", *Qualitative Inquiry*, Vol. 12 No. 2, pp. 219–245.

German Federal Government (2020), *Strategie Künstliche Intelligenz der Bundesregierung. Fortschreibung 2020.*, available at: https://www.ki-strategie-deutschland.de/files/downloads/201201_Fortschreibung_KI-Strategie.pdf (accessed 28 October 2022).

Glaser, B.G. and Strauss, A.L. (1967), *The discovery of grounded theory: Strategies for qualitative research*, *Observations*, Aldine, New York, NY.

Hölscher, I., Opiela, N., Tiemann, J., Gumz, J.D., Goldacker, G., Thapa, B. and Weber, M. (2021), "Deutschland-Index der Digitalisierung 2021. Kompetenzzentrum Öffentliche IT", available at: https://www.oeffentliche-it.de/documents/10181/14412/Deutschland-Index+der+Digitalisierung+2021.

McDermid, J.A., Jia, Y., Porter, Z. and Habli, I. (2021), "Artificial intelligence explainability: the technical and ethical dimensions", *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, Vol. 379 No. 2207, p. 20200363.

Mukhamediev, R.I., Popova, Y., Kuchin, Y., Zaitseva, E., Kalimoldayev, A., Symagulov, A., Levashenko, V., Abdoldina, F., Gopejenko, V., Yakunin, K., Muhamedijeva, E. and Yelis, M. (2022), "Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges", *Mathematics*, Vol. 10 No. 15, p. 2552.

Sovrano, F., Sapienza, S., Palmirani, M. and Vitali, F. (2021), "A Survey on Methods and Metrics for the Assessment of Explainability Under the Proposed AI Act", 8-10 December, Vilnius, Lithuania.

Sovrano, F., Sapienza, S., Palmirani, M. and Vitali, F. (2022), "Metrics, Explainability and the European AI Act Proposal", *J*, Vol. 5 No. 1, pp. 126–138.

Sovrano, F. and Vitali, F. (2023), *An Objective Metric for Explainable AI: How and Why to Estimate the Degree of Explainability*.

Strauss, A.L. and Corbin, J.M. (1990), *Basics of qualitative research: Grounded theory procedures and techniques*, Sage Publications, Inc.

Walke, F. and Winkler, T.J. (2022), "The TIHP Framework – An Instrument for Measuring Quality of Hybrid Services".

Walke, F., Winkler, T.J. and & Le, M. (2023), "Success of Digital Identity Infrastructure: A Grounded Model of eID Evolution Success.", *Proceedings of the 56th Hawaii International Conference on System Sciences*, Vol. 56.

Wiesche, M., Jurisch, M., Yetton, P. and Krcmar, H. (2017), "Grounded Theory Methodology in Information Systems Research", *Management Information Systems Quarterly*, Vol. 41 No. 3, pp. 685–701.

Williams, M. and Moser, T. (2019), "The Art of Coding and Thematic Exploration in Qualitative Research", *International Management Review*, Vol. 15 No. 1, pp. 45–55.