Forensic Science International: Digital Investigation 46 (2023) 301617

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

Digital forensics and strong AI: A structured literature review

Johannes Fähndrich [a,*], Wilfried Honekamp [b], Roman Povalej [c], Heiko Rittelmeier [d], Silvio Berner [e], Dirk Labudde [f]

[a] Hochschule für Polizei Baden-Württemberg, Sturmbühlstraße 250, Villingen-Schwenningen, 78054, Baden-Württemberg, Germany
[b] German Police University (DHPol), Zum Roten Berge 18-24, Münster, 48165, North Rhine-Westphalia, Germany
[c] Police Academy of Lower Saxony, Gimter Str. 10, Hann. Münden, 34346, Lower Saxony, Germany
[d] Central Office for Information Technology in the Security Sector (ZITiS), Zamdorfer Straße 88, Munich, 81677, Bavaria, Germany
[e] University of Applied Police Sciences Saxony, Friedensstraße 120, Rothenburg/OL, 02929, Saxony, Germany
[f] University of Applied Sciences Mittweida, Technikumplatz 17, Mittweida, 09648, Saxony, Germany

ARTICLE INFO

MSC:
68T01
68T99

Keywords:
Strong AI
Digital forensics
Artificial intelligence
Digital investigations

ABSTRACT

Forensics is an established field of research. Digital forensics started 44 years ago with the Florida Computer Crimes Act (1978) including legislation against the unauthorized modification of data on a computer system. Since then, the field has flourished in different subdomains. The overall definitions and concepts have been specified by a small group of experts. Furthermore, the need for development is created by the amount of digital evidence which is collected concerning most crimes. This paper gives an overview of the state-of-the-art by presenting a structured literature review of digital forensic about methods and concepts using Artificial Intelligence (AI) technologies. The review focuses on science done on topics in strong AI and forensics.

1. On the topic of digital investigations

With the increase of digitalization and the pervasiveness of information systems, a crime scene is no longer what it used to be with its mix of a location, people, evidence, changes in time, and their virtual counterparts. With the legislation against the unauthorized modification of data (Casey, 2004) evermore evidence including the mainstream use of smart homes, infrastructure, factories, or cities, investigations, and forensic evidence is no longer bound by one physical location.

The number of digital evidence has been increasing massively for years, and the large number can only be evaluated to a limited extent by human forensic specialists (Spranger et al., 2016). With the growing amount of digital information, an application of Artificial Intelligence (AI) in forensics is incumbent. The field of AI research and application has flourished (Jeong, 2020). Methods from Machine Learning and Data Science need to be extended to be explainable and valid for legal purposes. We have the goal of collecting work on strong AI with the application on forensic science, with the focus on properties of AI which are relevant to digital forensics. There are many surveys, collecting applications of methods of AI addressing problems in forensics. Thus, we chose another focus in this review. This literature review takes a more abstract view, not on the application of AI in forensics, but in the publication which combines the abstract idea of AI with forensics. In this survey we create an overview on top level domains which are AI and digital forensics. We do not include applications of AI in narrow domains.

Thus, the research questions analyzed in this survey is: "**Is there research on strong AI topics concerning digital forensics?**"

This is done to shed light on the properties of the methods of AI like explainability, which could help the broad introduction of AI in the field of forensics. We hope to launch a new investigation of AI researchers on the fundamental properties wished for with the application of AI in digital forensics. Those properties are mostly fundamental challenges of computer science, and they depend on the used AI models. Therefore, we want to initiate scientific work on those properties. We neglect the issue concerning accuracy or calculation efficiency, since this is mostly evaluated before the application of methods of AI to a topic. There are many challenges, but some of them are prioritized. We will take a deeper look at some of them in regard to terminology, next:

Explainability (Sanchez et al., 2019) As in forensics, one goal of data analysis is to create evidence to be used in court, non-

* Corresponding author.
E-mail address: johannesfaehndrich@hfpol-bw.de (J. Fähndrich).

https://doi.org/10.1016/j.fsidi.2023.301617
Received 7 October 2022; Received in revised form 24 July 2023; Accepted 4 August 2023
Available online 25 August 2023

experts are confronted with the results. To understand the evidence created by methods of AI, the way the used algorithms or models needs to be transparent and explained. By showing the inner working of the used method, it could be shown how a result has been reached.

Human in the loop (Nguyen and Choo, 2021) Most Methods of AI do not work on a level of accuracy that is acceptable for forensic investigations. One method of coping with errors is to take a human back into the process and supervise the result of the algorithm. Some models allow a confidence estimate of the output. The confidence indication on the result reduces the cases in which the human has to be involved. Black-Box Models often need additional steps to create output confidence intervals (Guo et al., 2017). But nevertheless, including a human in the process enables the system to handle the faults of methods of AI in multiple ways: First errors can be seen and censored, before making a decision based on the output of the system. Second, new training data can be annotated. With that, the system can improve with each error it makes. Third, parameters can be adapted to specific problems and the model does not need to perform well on all tasks, e.g. the language selection before automated translations. This enables the system to train different models for different languages, and the selection of the best fitting model is done by the human.

Adversarial models (Nowroozi et al., 2021) is an intrinsic problem of AI models: If a model is trained and not a general Artificial Intelligence, we can train an adversarial model, which creates input to the model, producing errors. Depending on the model and the context, there are different types of creating an adversarial model, which creates a precalculated or random output of the AI model (Zhang et al., 2020). The existence of adversarial models is an argument against the use of AI in forensics in general.

Bias in data and models (Meissner and Kassin, 2002) reducing bias in data sets used by AI and the resulting models has been in the focus of ethical AI research (Raji et al., 2020). The generated bias should be reported with each result used in forensics evidence.

Model sharing (Veale et al., 2018) is the idea that models are trained by one party and shared with others to reduce training effort. By sharing the model, the model can be subject to attacks. One outcome can be that private data can be extracted from the model. In addition, sharing a model allows for white-box adversarial models.

Autonomy (Totschnig, 2020) is the property of a system to make decisions on its own. To make a decision autonomously, one assumption is that there is the freedom to decide between options. Which is leading to the risk to select a suboptimal option. With an autonomous decision, the question of explainability becomes more complex, since the values on which the decision is based, might not be purely data driven but could be contextual.

Consciousness and conscience (Meissner, 2020) as in the ability of being self-aware and embedding values in one's own belief system, needs to be discussed if autonomous decisions are made by AI systems. Meissner added a hierarchy of such properties, which is depicted in Fig. 1 and should be discussed.

We added to each step a classification as done in this review and its downside to forensics, e.g. the explainability property of machine learning approaches is part of the first level. Since the results of a machine learning model are hard to understand and if a model creates documentation on how it reached its decision, it becomes easier to understand. But because these are mostly statistical models learned from data, they can be explained. This is questionable in the second level, where the models do not have to be statistical and become more complex. Depending on the elaboration of the intelligence, it might no longer be possible for humans to understand the reasoning or complexity of information that leads to a decision. From there the properties are more abstract and fuzzy and with that harder to classify. Creativity for example is used to create solutions to new problems. The evaluation of such a non-standardized process could, e.g. be more effort on analyzing if it stands up in court. Finally, conscience leads to further problems like goal attainment and management. Since autonomous, self-aware systems would probably manage their own goal, which could conflict with those of an investigation.

Trust (Marcus and Davis, 2019; Siau and Wang, 2018) in the ability of AI and its results is one of the more abstract solutions. The problem at hand here is: if AI becomes sufficient sophisticated, human intelligence is no longer adequate to understand the reasons of such a system. Thus, we need to resort to trusting the results without being able to check them.

These properties are just examples of abstract properties of AI that could be discussed. During our investigation in this survey, we noticed that most papers are concerned with weak AI, applying machine learning and data science to problems in forensics (Hall et al., 2021; Jarrett and Choo, 2021). In more details, we classified the different topics between the first two layers in Fig. 1 in more detail in Fig. 2.

As the pace of AI research continues to accelerate, the gap between the state of the art and its application in forensics is widening.

2. The state of decay

With accelerated technological development, the expert level of law enforcement trails this development with the needed insight to handle modern digital forensics. New versions of operating systems, like company-specific Android clones from Sony, Huawei, or Samsung, make it even harder to build up expertise and good protocols for forensic investigations. On top of that, we have a massive amount of apps developed and updated weekly. Getting to know all the possible digital evidence, needs additional research and without any abstraction is a Sisyphus task. The last but most discussed challenge is the amount of data created and collected by law enforcement. With an ever-growing capacity of storage in consumer electronics, a simple case of a bar brawl with criminal assault concludes in the seizure of multiple smartphones, and the need to analyze gigabytes of data.

With the speed of technological development, the knowledge taught in the education of law enforcement officers in digital forensics decays every day. The gap between the state-of-the-art in the negative dual-use potential of modern IT-Systems and methods of Artificial Intelligence, and the knowledge of investigators seems to grow constantly. This conclusion has been drawn from the structured litterateur review (Armitage and Keeble-Allen, 2008), which is presented in section 3. The dual use potential of AI has become a subject of research in many domains (Schmid et al., 2022; Urbina et al., 2022). Even politics has begun to engage to AI and the dilemma of dual-use (Kania, 2018). We see three types of results coming from this discussion: The abandonment of AI Research, the use for crime and the utilization in domains to benefit like digital forensics.

Machine Learning (ML) has become a major technology for many industries. The improvements in the methods of automating statistical analysis have made today's ML algorithms outperforming humans in many tasks. Complex models like Neural Networks so big they are called Deep Learning are at the forefront of this scientific endeavor. With Deep Learning, we tackle ever more complex task, which leads to more complex models. The complexity of the models has become incomprehensible or better unintelligible for most humans. This means we get results from the application of ML, but might no longer under-

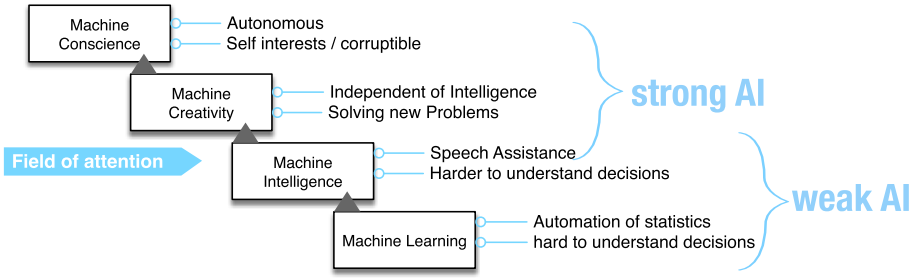

Fig. 1. Interpretation of the hierarchy of abstract properties of AI (Meissner, 2020).

stand the created solution. Explainable AI (XAI) tries to shed light on the inner working of such complex models used in modern ML Algorithms (Samek et al., 2019). XAI, therefore, is part of the research of AI and ML algorithms, to make the functional principles of the methods more understandable for humans.

"Despite all these developments, the promises of strong artificial intelligence set forth in the 1960s have not been fulfilled." (De Winter and Dodou, 2014) [p. 7] The automation of statistics as done in ML is not yet general intelligent behavior. Small but complex problems can be solved with ML up until today. But topics like cognition, planning, learning, reasoning and pragmatic language understanding are part of the broader research are call Artificial Intelligence (Ertel, 2018). Argument of the feasibility of AI in the broader sense has been made (Cole, 1991). This AI in a broader sens is called "strong AI". Strong AI has been defined as antagonism to "weak AI", which describes ML as it is practiced today. But it can be argued that Machine Intelligence, and with that ML, is part of strong AI. For strong AI to be applied to digital forensics, strong AI topics have to be further research since common understanding in AI research is, that strong AI has not yet been developed (Nordby et al., 2022).

3. Structured literature review

Artificial intelligence has reached many domains. The idea of not programming algorithms to solve a problem, but to collect data and let machine learning create a solution, looks promising. Although this is not AI, it is the application of methods of AI (in this example the machine learning part) to a domain-specific problem. There are two ways AI can be involved in crimes, either as a tool or as a target (Jeong, 2020).

AI has been a topic of research since 1950 when Turing asked the question "Can a machine think?" Artificial Intelligence as a term has been coined in 1955 by McCarthy et al. (1955). Till then, the research field has grown and specialized in many subdomains. The research can be structured into two goals (Flowers, 2019): Weak AI and Strong AI. The first utilizes methods of AI to solve problems like the ever-growing amount of data collected as evidence. The second describes aims at producing a system comparable to a mind including intelligence, cognition, understanding and other mental states.

Many surveys have analyzed the use of weak AI application to forensics. Neglecting the questions of strong AI and its implications for the domain of digital forensics. Therefore, this survey attempts to locate contributions which look at AI and its properties and its applicability in digital forensics. With the question in mind: What needs to change on how AI works, to be applicable[1] in forensics?

The application and the questions of the validity of methods in research in weak AI are direct statistical evaluations. Statistical evaluation is done in experiments, or applications in a domain and the result in its usefulness is calculated (e.g. through F1-scores, confidence intervals or ANOVAs). In the domain of strong AI, such an evaluation is harder, since e.g. the notion of conscience is less formally described. There are many publications on the application of methods of AI to special topics, e.g. Hoelz et al. (2009); Mitchell (2010); Jeong (2020); Veldhuis et al. (2022), and they also contributed to digital forensics. But the discussion of topics of strong AI in critical applications[2] is less researched, and we want to emphasize research in this area. Current literature reviews are mostly specific, e.g. Ganesh et al. (2022); McKinnel et al. (2019); de Sousa et al. (2019). Therefore, this survey is directed to the analysis of papers actually naming Artificial Intelligence and digital forensics in the title.

In both domains, topics overlap and some of them are researched by both types of research, but on different abstraction levels. Surveying the depth of each field of AI is a task that would take up too much space for this journal. The idea of this review is to survey publications that analyze the application of AI methods of both types to digital forensics. With the classification shown in Fig. 2 we analyze the topmost terminology AI and not its implementation in Machine Learning methods or further down the specialization of the application of Data Science in digital forensics.

We have chosen those keywords because we want to analyze the use of strong AI in the domain of digital forensics and law enforcement, and not the application of a specific AI method like machine learning (weak AI) in the domain of digital forensics.

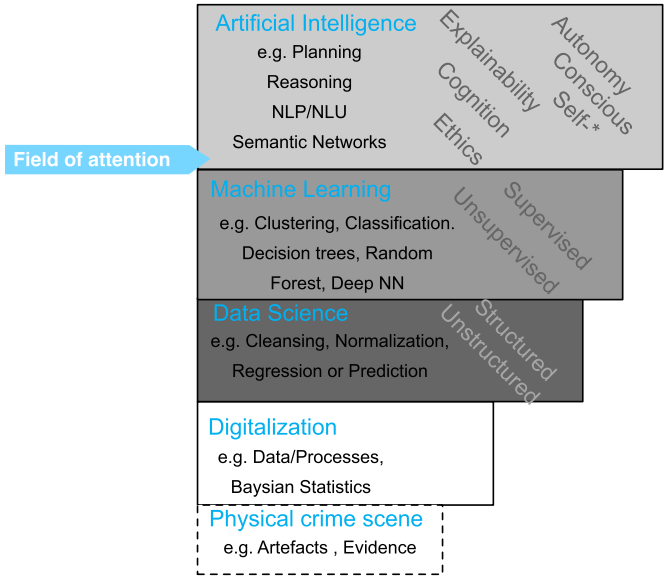

Fig. 2. Embedding of Topics of AI and related terms (own illustration).

We create a literature review by looking at the following resources using the PRISMA 2020 Workflow (Page et al., 2021), although not all

[1] e.g. explainable or trustworthy enough.

[2] critical meaning impacting humans lives deeply.

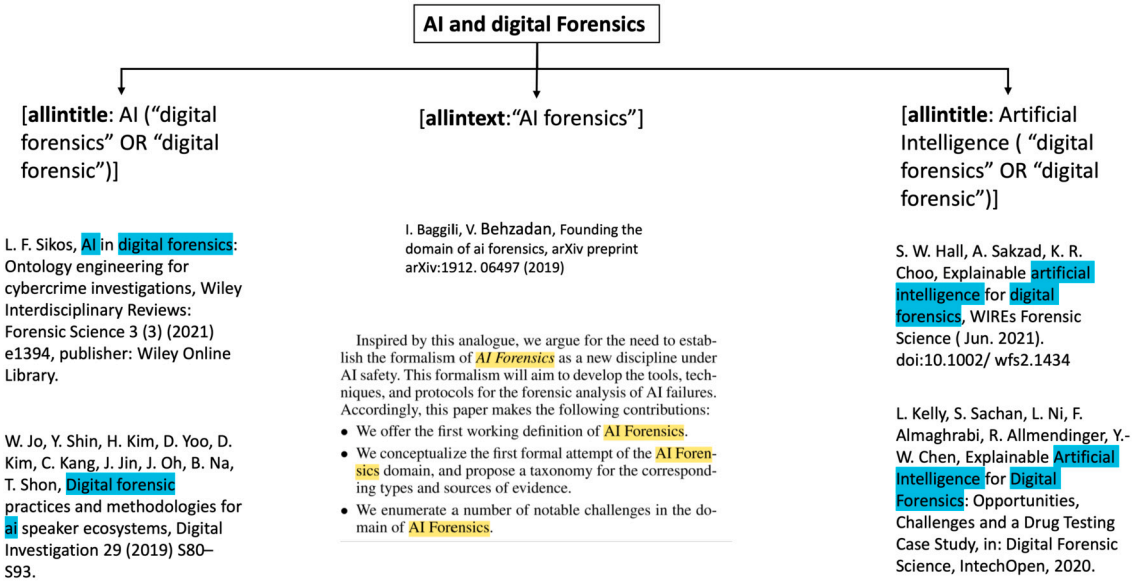

Fig. 3. Overview of Methodology.

steps of the checklist are applicable. Since we used expert reviews of the papers, e.g. bias analysis and synthesis methods are too heterogeneous to describe. Nevertheless, we structured the method and results as much as possible (Fig. 3).

In the first step we used Google Scholar with search terms **AI digital forensics** and **Artificial Intelligence digital forensics** to identify fitting papers. Then, we scrutinized the more than 10,000 results and derived search terms to narrow down the research (Schmid et al., 2022). Thus, we finally combined ("digital forensics" OR "digital forensic") with "AI" as well as "Artificial Intelligence" and searched for this in the title of the publications. Additionally, we discovered a new domain called "AI forensics" and searched for this term in the text. From the results, we removed duplicates and other unscientific publications. Excluded were, e.g. papers published in journals or conferences which have been listed as predatory publishers.[3] Furthermore, we removed presentations, student project reports, Bachelor theses and nonscientific publications. Additionally, we removed publications we could only locate on arxiv.org or EasyChair preprints because we can not grantee that the papers have gone through a peer review. Results were limited to the last five years (since 2017). The papers have been analyzed by the authors due to resources, only one scientist did examine an individual paper. The following research led to the work presented here.

Search string [allintitle: AI ("digital forensics" OR "digital forensic")]

The short paper by Constantini, Lisi and Olivieri describes a research network infrastructure that brings digital forensics experts AI researchers together (Costantini et al., 2019b). This network is supposed to foster scientific advances in the area of artificial intelligence for digital forensics. It has been established in the frame of the EU-funded project digital forensics: analysis tests through intelligent systems and practices (DigForASP). The authors expect an impact on the development of AI applications for analysis of digital evidences and decision support systems, as well as increased transparency.

The paper by Sikos (2021) describes the challenges of digital forensics and the limitations of the state-of-the-art. The main contribution of this paper is the suggestion to describe the domain of digital forensics in RDF and OWL.

The result by Cruz (2019a) is a master's thesis that met the criteria of this structured literature review. Since it has been reviewed by two professors, we will see this as peer-reviewed and analyze it in this paper. Cruz describes artificial intelligence in two applications: Machine Learning and Natural Language Processing (NLP). This is quite narrow and shows that weak AI is used in this work. Furthermore, he proceeds in the application of crime detection on online data. This explains the focus on machine learning and NLP, since online most information is in text form, and information can be crawled to be used as a data set and an input for methods of machine learning. Cruz (2019a) proceeds with two examples of how methods of AI were analyzed in lawsuits in US courts. Unfortunately, this thesis stops before a real experiment is conducted.

Kim et al. (2021b) conducted a study of user data extracted from wearable devices. The term artificial intelligence used in the title of the paper and the abstract only refers to the devices and has no respect to the way digital forensics is applied.

Jo et al. (2019) performed digital forensic analyzes of four AI speaker ecosystems. They derived and proposed five digital forensic analysis methods and practices for those. Neither weak nor strong AI are used for their analysis. In this respect, the authors' paper is outside the focus of this review.

Jang and Shin (2021) proposed in their paper an AI-based evidence collection system. The basic idea of such a system is well described. However, there is no deeper explanation about concept, requirements, design, development, testing, deployment etc., neither in relation to artificial intelligence nor to digital forensics.

The "Special Issue on Application of AI in Digital Forensics" (Fähndrich et al., 2022) is a collection of papers on AI with application to forensics, focusing on the fusion of computer science, data analytics, and machine learning with discussion of law and ethics for their application to cyberforensics. It includes technical contributions, a discussion paper, system description, project report, and an interview with one of the leading experts in the science of AI. The authors explain in the introduction, that the chain of custody, the legal certainty and the data protection are major hurdles for the use of AI. Therefore, in order to be able to cope with the ever-increasing

[3] compared with https://beallslist.net.

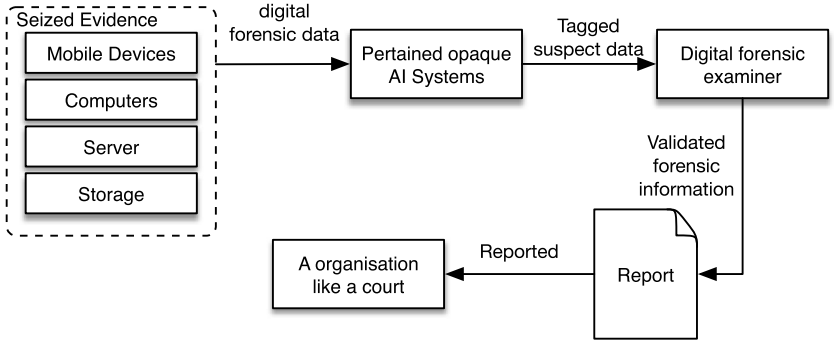

Fig. 4. State-of-the-art of the use of methods of AI in a forensic process (based on Hall et al., 2021).

number of potential sources of evidence, machine learning and data science methods must be extended to be explainable and valid for legal purposes.

Search string [allintitle: Artificial Intelligence ("digital forensics" OR "digital forensic")]

We were unable to find the paper of Sanyasi and Kumar (2021). The only published contribution is an abstract, which makes it difficult to evaluate this contribution.

Hall et al. (2021) describe their publication as an opinion letter, where they discuss the application of Explainable AI (XAI) to digital forensics. The state-of-the-art is described in a figure much like Fig. 4

The methods of AI are here utilized to automate minor tasks like the detection of drugs, weapons, categorize chats and images. In their outlook, Hall et al. postulate that topics like data discovery and recovery, device triage, network traffic analysis, encrypted data forensics, timeline/event reconstruction, and multimedia forensics might profit from the integration of methods of AI into their fields of research.

Costantini et al. (2019a) propose the use of Answer Set Programming as a logical computational tool for the support of digital forensics. With their analysis of the state-of-the-art, they conclude, However, support for the effective aggregation and organization of useful evidence is simply non-existent. They aim at a Decision Support Systems (DSS) for future investigators, which can help analyzing the collected data. This is because according to current legislation, they can merely be auxiliary tools, and not substitutes, to the human decision-maker. This means that we still have a way to go for the AI systems to prove their lack of bias, their consistency, or their ability to aid law enforcement in their investigations. Costantini et al. (2019a) describe examples of how ASP can be used to reason in different data types. Whether this reasoning will hold up in court remains to be seen.

Kelly et al. (2020) discuss the use of AI in digital forensics and the need of utilizing explainable methods. They ask questions about the models used, which help classify an AI method, and discuss their answers, e.g. intrinsic vs. post-hoc explanations, model-specific vs. model-agnostic, or local vs. global. Kelly et al. analyze additional obstacles, which encounter with uncleaned forensic data like missing date, feature selection or interactive learning.

Fig. 5 shows one of the different perspectives to processes in digital forensics. The one shown by Kelly et al. is an abstract, which represents a view of a data scientist. Data is cleaned and normalized. Machine models are learned and selected depending on the data and the context of the classification or clustering task. The models are trained and tested and after a quality thresh-hold has been reached the models are deployed for their use.

Rughani (2017) analyzes frameworks, which are used in digital investigations. The process is separated into the three classical steps: acquisition, analysis, and presentation. The analysis does not deeply describe the usage of AI methods in the tree steps, but shows possible applications. With that, the contribution is only an idea without evaluation or implementation.

Jarrett and Choo (2021) see the potential of the application of AI to digital forensics, with AI as "massive scale to automate a wide variety of processes and operations". Here, Jarrett and Choo understand AI as a tool to automated task in digital forensics. According to their analysis of complaints processed in the FBI IC3 from 2001 to 2019, the total number of internet crime complaints grew from approximately 50,000 in 2001 to about 467,000 in 2019. Total damages by internet crime grew from 17.8 million to 3.5 billion Dollars. This means the overall damage by reported internet crime has become almost 200-times higher. Jarrett and Choo (2021) classify the terms Artificial Intelligence, Machine learning and Deep Learning as means to automation. They conduct a literature review on AI application on automation, then the application of automation in digital forensics, and with that finally the application of AI in digital forensics. They focus on Tools and Frameworks of AI. As a motivation, Jarrett and Choo (2021) determine that the average cost of an expert forensic examination ranges from $5,000 to $15,000 and in complex cases it can exceed $100,000. Furthermore, they argue that the speed of digital investigation can be improved by methods of AI. But the cost and speed of a digital investigation is just partially the motivation of using AI in digital forensics: We argue that clues are missed due to the amount of data that needs to be investigated. Thus, the use of AI does not only lower the cost of digital investigation, but also increases the investigation result quality.

Iqbal et al. (2020) describe in their chapter applications of AI methods in different Domains. The contribution of this chapter is an analysis of software used by law enforcement utilizing data mining and machine learning. Those are, e.g., Chen et al. (2003) or Nissan (2012). The analysis of those tools shows, that there has been the first application of methods of AI into forensic tools. The analyzed examples are Deep Learning in Social Media Mining like Tweet Crawler for Events Identification or Information extraction. Their analysis concludes in future applications which are manifolds, and they conclude The implementation of AI holds the potential to dramatically change the field of digital forensics (Iqbal et al., 2020, p. 149).

Raponi et al. (2022) describe a special system for shot detection of gunshots. In the title they use Digital Forensics and Artificial Intelligence, therefore they are reviewed in this structured literature review, even though this paper is the

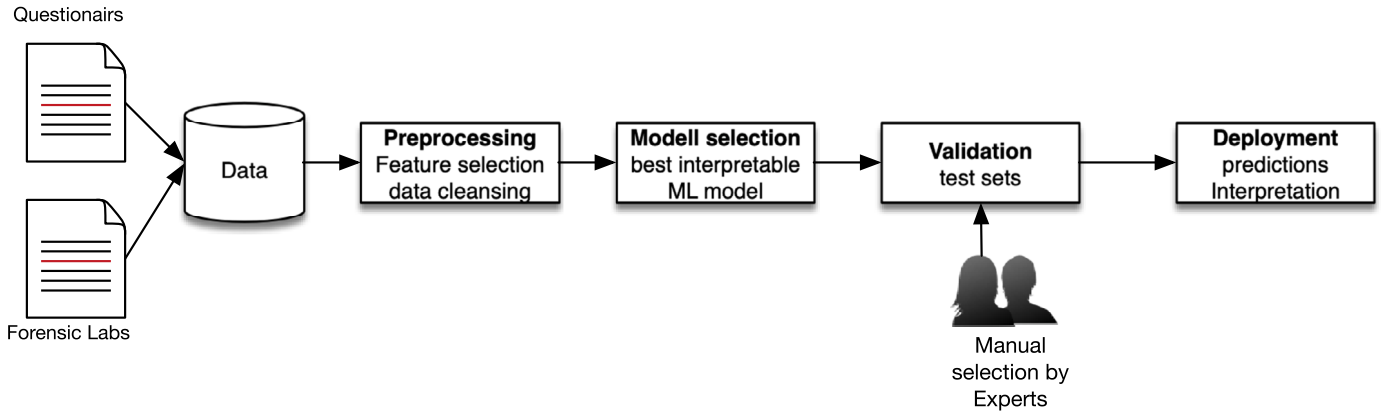

Fig. 5. One machine learning perspective of a forensic process (Kelly et al., 2020).

application of narrow AI to a specific problem. Here, a standard of machine learning (a Support vector machine) is used to train on a training set to locate and identify different gun types by the sound they make if a shot is made. Even though this meets the criteria, this is not a paper we want to analyze in this survey. But it is a good example of the state-of-the-art on narrow AI.

Malhotra (2023) is a book chapter. They argue for the use of AI in digital forensics, but with the formal description of the used processes, data. This formal description is needed to make evidence admissible in a court of law. Thus strengthening our argument for the analysis of Methods of AI in the topic of strong AI.

Cruz (2019b) describes in his Master's thesis the effectiveness of AI in digital forensics. Machine Learning and Natural Language Processing are classified as implementations of AI. The focus here is on the detection of online crime and its disclosure in court.

Chen (2020) proposes an architecture for AI-based examination of digital forensic evidence where artificial intelligence systems are paired with human experts. Human investigators are responsible and accountable for the accuracy of the investigation. Machines are responsible for speed, humans for accuracy.

Search string [allintext:"AI forensics"]

Baggili and Behzadan (2019) in their conference paper coined the term "AI Forensics" for a new discipline subordinate to AI safety that aims at the development of tools, techniques and protocols for the forensic analysis of AI failures. They "propose a taxonomy for the corresponding types and sources of evidence" and consider explainability as one of the main challenges of AI in forensics.

Faraldo Cabana (2023) is a book chapter. It describes the hurdles which have to be taken by an example application of machine learning for investigation. Taken from other forensic identification methods, the idea of arguing with examples is a good one and could be a valid approach to the legal problems. The chapter discusses a special application, but the argument could be generalized.

Kim et al. (2021a) describe a browser extension using ML to classify malware. The AI Model is used to identify websites with malicious software. This is an example of weak AI. The term forensics is used to describe the browser extension capability to analyze the browser caches. The contribution is seen as a proof-of-concept implementation of "applying AI technology to digital forensics".

Atlam et al. (2020) create a state-of-the-art analysis to the study of forensics for the Internet of Things (IoT). The analysis extracts requirements, challenges, solutions and open research directions for forensics in the IoT. The need of automation (via AI) is discussed.

Bhatt and Rughani (2017) review how machine learning can be used in digital forensics. The paper is not scientific and describes a tool in an unstructured manner. ML an AI is seen in an implementation of a DeepQA.

Solanke (2022) gets to the heart of the problem with the focus of explainability. The paper describes different aspects of explainability and classifies possible solutions into "white box interpretable models" and "Post-hoc explanations". The main contribution of the paper is the crystallization of recommendations to gain trust in models. Thus interpreting explanations as mitigating for trust in the performance of ML models. The discussion in Kim et al. (2021a) shows how complex the discussion of ML models and trust is. The subsumption into strong AI topic is still needed, but the analysis can be built upon.

Brighi et al. (2020) in their workshop paper address the question of how the processing time and the reliability of the results in the analysis of digital evidence can be supported by the use of AI. Therefore, they show the connections between the principles of forensic investigation and AI applications. They state that in order to define the legal framework for AI applications in digital forensics, their functioning must be fully understood. Furthermore, the boundaries between legally acceptable and unacceptable consequences must be determined. They support the development of monitored systems in which interpretability results from the use of humans and call this humanware in the field of digital forensics.

Krishnan et al. (2022) aimed to quicken the review and analysis phase through developing a custom forensic software that automates the handling of case evidence and leverages analytics to predict sentiments of case suspects, indicators of financial fraud and sexual harassment of suspects while pointing to their evidence sources. The author highlighted in his study, that in the case of large volumes of data, automation coupled with data mining and artificial intelligence can greatly speed up the forensics process and thereby allow for a quicker investigation. Further on he explained, that machine learning, convolutional neural networks, and natural language processing are having a high impact on electronic discovery and forensic investigations. Particularly making decisions using artificial intelligence needs to be continuously explainable to a jury. For this reason explainable artificial intelligence will need to be employed.

The focus of the work of Leone (2021) is on the application of recognition algorithms using AI. Not the direct

application is considered, but the discussion in the field of semiotics. Semiotics defines humans not as objects, but as a matrix of signs. These signs, according to the paper, cannot always be recognized by AI and need to be critically discussed in this field. On the basis of cultural patterns this is explained in detail. Interesting are the remarks on today's tendency of "naturalization" of technology and the resulting digital imaging techniques.

This monograph was prepared as part of a dissertation of Horan (2022). The focus of the work is on the development of a framework for the automated collection of information using the approaches of OSINT. OSINT (Open-Source-Intelligence) provides a lot of information with different tools, which are found in unrelated reports. The developed framework automates the information collection, summarizes these reports and creates a single graphical report. Machine learning approaches are used as part of the automation.

In the paper of Edwards et al. (2022), an expansion of the application area of AI forensics takes place. AI Forensics is a novel research field that aims at providing techniques, mechanisms, processes, and protocols for an AI failure investigation. The goal of this work is to establish a subfield of AI forensics, namely AI model forensics. AI model forensics examines the forensic investigation process, including where available evidence can be collected. Starting with the nature of the development and use of Ki models, we explain that these models can be replaced, infected, or fooled by adversarial input. Using a literature review, we examine the relationships and dependencies of AI model forensics with the subfields of software forensics, cloud forensics, and network forensics, med that are useful for AI investigations. Overview of digital forensics practices in areas strongly related to AI and useful for implementing protocols and procedures specific to AI investigations. The potential applicability of the field of AI model forensics is explained using a scenario (AI-based model decides on surgery) and provides a perspective on the benefits and challenges.

The method of review of the papers has been a qualitative study by one of the authors as an expert in digital forensics of the selected papers, assessing and summarizing the main findings of each paper looked at. The papers have been analyzed by one of the authors as specialist in digital forensics and computer science. With that, we hope to spark confidence in the reader in our conclusion.

4. Discussion

Restricting the structured literature review to the keywords search in the title filters most of the specific papers out of the analysis. The results show the usage of the term AI and forensics in abstract fields, and not the application of one type of, e.g., machine learning methods in forensics. A broader search on the application of methods of data science, machine learning, and even AI results in more papers but neglects the abstract questions of applicability of AI in digital forensics. Broadening the search of methods of AI in the application can be seen as an investigation of weak AI in digital forensics. The analysis of this survey is to show that a discussion is needed, on explainable and trusted AI.

Results from the analysis are, that the application of AI in digital forensics is mostly restricted to the application of different machine learning methods to narrow problems in a specific domain. The bigger issues like the questions about trusting autonomous systems in their decisions, a discussion of adaptions of laws, or forensic implications of the automated analysis of data not available today, are still missing. The application of machine learning to the available data of a certain context like a city is an approach studied to start the evaluating methods of AI (Shapiro, 2017). The development possibilities are many-fold, a detection of discriminate patterns in police work (Brantingham et al., 2018), the Digitalization and analysis of cold cases, or the live selection of data gathered are just some visions.

Given the speed at which AI is evolving (Liu et al., 2021), every day missed means another day of lag for law enforcement. Every missed opportunity to explore an application of data analytics, machine learning or methodes of AI puts the application in a negative dual-use potential and leaves law enforcement's tech affinity in decline.

The analysis of weak AI in any application does not mean that the methods include decision-making, autonomy, or intelligence. Meaning, humans are still making the real decisions and the methods of AI are only decision support systems. Thus, its application is not as critical as the application of strong AI. Which in its definition is autonomous and with that makes its own decisions. Thus, more analysis is needed on the trust in the results of strong AI systems, especially in the application in law enforcement. Even though the development of Methods of AI which merits the label strong AI is still ongoing, more and more automation in digital forensics is needed, and with that the questions of trust in those methods need to be raised.

For scientists, we want to motivate the research on topics placed in the areas of Strong AI. This survey shows the gap between the use of weak AI in many domains, which are established in the field of forensics, and the topics of strong AI, which are not yet part of the scientific disscourse. The research on these topics could make it easier to use results created by AI systems before a court.

The properties of processes in digital forensics and forensic readiness have been subject to research (Rowlingson et al., 2004; Pan and Batten, 2005; McKemmish, 2008; Damshenas et al., 2014; Kebande and Venter, 2018; Amato et al., 2020) (chronological order). This survey has shown that connecting forensic requirements and methods of AI (even weak once) is still staring out. For the even more difficult problem of strong AI, the discussion has not been started yet.

5. Conclusion

Ten years ago, Simson Garfinkel predicted the state-of-the-art (Garfinkel, 2010): Increasingly organizations encounter data that cannot be analyzed with today's tools because of format incompatibilities, encryption, or simply a lack of training. Even data that can be analyzed can wait weeks or months before review because of data management issues. Without a clear research agenda aimed at dramatically improving the efficiency of both our tools and our research process, our hard-won capabilities will be degraded and eventually lost in the coming years. In Germany this is the case: evidence takes months to be processed, encryption lets surveillance "going dark" and training is miles away from being sufficient to narrow the gap between the knowledge of law enforcement and the technologies used.

The takeaway message from this survey can be: we need more application of AI in digital forensics. And with that, research concerning properties which help the usability of results of the application of methods of AI to forensic problems. To conclude: we need more fundamental research on AI to be able to apply it in domains like digital forensics. Explainability, robustness to bias and accuracy are properties which need more focus if the fundamental questions of strong AI (like shown in the outermost circle of Fig. 2) are to be addressed. A good example of the lack of such research could be seen in autonomy and consciousness: The research in this study did not produce any papers with a discussion of autonomy or conscious AI in digital forensics. As this can easily be turned into a fundamental question of AI research, its application in law enforcement might be of interest.

For future work, the search terms of a survey could be broadened: "Machine Learning" as well as "Data Science" could yield more specific results on which methods of AI are currently applied to digital forensics and which need further investigation. The transfer of research results from different research areas of AI has to be flanked by evaluating a

moral, ethical and law perspectives, while evaluating their application in the domain of digital forensics.

Future research could be directed in asking the questions of strong AI to be discussed for the use in our justice systems: How do we use systems, we do not fully understand? Which decisions are we willing to let AI decide? How do we integrate autonomous systems into trails? How do we establish protocols if machines disagree in their output given the same data?

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Amato, F., Castiglione, A., Cozzolino, G., Narducci, F., 2020. A semantic-based methodology for digital forensics analysis. J. Parallel Distrib. Comput. 138, 172–177.

Armitage, A., Keeble-Allen, D., 2008. Undertaking a structured literature review or structuring a literature review: tales from the field. In: Proceedings of the 7th European Conference on Research Methodology for Business and Management Studies: ECRM2008. Regent's College, London, p. 35.

Atlam, H.F., Hemdan, E.E.-D., Alenezi, A., Alassafi, M.O., Wills, G.B., 2020. Internet of things forensics: a review. Int. Things 11, 100220.

Baggili, I., Behzadan, V., 2019. Founding the domain of ai forensics. arXiv preprint arXiv: 1912.06497.

Bhatt, P., Rughani, P.H., 2017. Machine learning forensics: a new branch of digital forensics. Int. J. Adv. Res. Comput. Sci. 8 (8).

Brantingham, P.J., Valasik, M., Mohler, G.O., 2018. Does predictive policing lead to biased arrests? Results from a randomized controlled trial. Stat. Public Policy 5 (1), 1–6.

Brighi, R., Ferrazzano, M., Summa, L., 2020. Legal issues in ai forensics: understanding the importance of humanware. In: On Applications of AI to Forensics 2020 (AI2Forensics 2020), p. 13.

Casey, E., 2004. Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet, 2nd edition. Academic Press, London, San Diego, Calif. oCLC, ocm53356563.

Chen, H., Zeng, D., Atabakhsh, H., Wyzga, W., Schroeder, J., 2003. Coplink: managing law enforcement data and knowledge. Commun. ACM 46 (1), 28–34.

Chen, J.Q., 2020. Ai-Enabled Digital Forensic Evidence Examination. Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), vol. 1. Springer, pp. 832–841.

Cole, D., 1991. Artificial intelligence and personal identity. Synthese 88, 399–417.

Costantini, S., De Gasperis, G., Olivieri, R., 2019a. Digital forensics and investigations meet artificial intelligence. Ann. Math. Artif. Intell. 86 (1), 193–229. publisher: Springer.

Costantini, S., Lisi, F.A., Olivieri, R., 2019b. Digforasp: a European cooperation network for logic-based ai in digital forensics. In: CILC, pp. 138–146.

Cruz, E.G.D.J., 2019a. The Effectiveness of Digital Forensics and Security Strategies in Using AI and Machine Learning to Protect Children Online. PhD Thesis, Polytechnic University of Puerto Rico.

Cruz, E.G.D.J., 2019b. The effectiveness of digital forensics and security strategies in using ai and machine learning to protect children online. Ph.D. thesis. Polytechnic University of Puerto Rico.

Damshenas, M., Dehghantanha, A., Mahmoud, R., 2014. A survey on digital forensics trends. Int. J. Cyber-Secur. Digit. Forensics 3 (4), 209–235.

de Sousa, W.G., de Melo, E.R.P., Bermejo, P.H.D.S., Farias, R.A.S., Gomes, A.O., 2019. How and where is artificial intelligence in the public sector going? A literature review and research agenda. Gov. Inf. Q. 36 (4), 101392.

De Winter, J.C., Dodou, D., 2014. Why the fitts list has persisted throughout the history of function allocation. Cogn. Technol. Work 16, 1–11.

Edwards, T., McCullough, S., Nassar, M., Baggili, I., 2022. On exploring the sub-domain of artificial intelligence (ai) model forensics. In: Digital Forensics and Cyber Crime: 12th EAI International Conference, ICDF2C 2021, Virtual Event. Singapore, December 6–9, 2021, Proceedings. Springer, pp. 35–51.

Ertel, W., 2018. Introduction to Artificial Intelligence. Springer.

Fähndrich, J., Honekamp, W., Povalej, R., Rittelmeier, H., Berner, S., 2022. Special issue on application of ai in digital forensics. Künstl. Intell., 1–4.

Faraldo Cabana, P., 2023. Technical and legal challenges of the use of automated facial recognition technologies for law enforcement and forensic purposes. In: Artificial Intelligence, Social Harms and Human Rights. Springer, pp. 35–54.

Flowers, J.C., 2019. Strong and weak AI: deweyan considerations. In: AAAI Spring Symposium: Towards Conscious AI Systems, pp. 1–7.

Ganesh, N., Venkatesh, N., Prasad, D., 2022. A systematic literature review on forensics in cloud, iot, ai & blockchain. Illum. Artif. Intell. Cybersecur. Forensics, 197–229.

Garfinkel, S.L., 2010. Digital forensics research: the next 10 years. Digit. Investig. 7, S64–S73.

Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On calibration of modern neural networks. In: International Conference on Machine Learning, PMLR, pp. 1321–1330.

Hall, S.W., Sakzad, A., Choo, K.R., 2021. Explainable artificial intelligence for digital forensics. WIREs Forensic Sci. https://doi.org/10.1002/wfs2.1434. https://onlinelibrary.wiley.com/doi/10.1002/wfs2.1434.

Hoelz, B.W., Ralha, C.G., Geeverghese, R., 2009. Artificial intelligence applied to computer forensics. In: Proceedings of the 2009 ACM Symposium on Applied Computing, pp. 883–888.

Horan, C., 2022. Open-source intelligence investigations: Development and application of efficient tools. Ph.D. thesis. University of Kansas.

Iqbal, F., Debbabi, M., Fung, B.C., 2020. Artificial Intelligence and Digital Forensics, in: Machine Learning for Authorship Attribution and Cyber Forensics. Springer, pp. 139–150.

Jang, E.-J., Shin, S.-J., 2021. Proposal of ai-based digital forensic evidence collecting system. Int. J. Internet Broadcast. Commun. 13 (3), 124–129.

Jarrett, A., Choo, K.R., 2021. The impact of automation and artificial intelligence on digital forensics. WIREs Forensic Sci. https://doi.org/10.1002/wfs2.1418. https://onlinelibrary.wiley.com/doi/10.1002/wfs2.1418.

Jeong, D., 2020. Artificial intelligence security threat, crime, and forensics: taxonomy and open issues. IEEE Access 8, 184560–184574.

Jo, W., Shin, Y., Kim, H., Yoo, D., Kim, D., Kang, C., Jin, J., Oh, J., Na, B., Shon, T., 2019. Digital forensic practices and methodologies for ai speaker ecosystems. Digit. Investig. 29, S80–S93.

Kania, E., 2018. Technological entanglement: Cooperation, competition and the dual-use dilemma in artificial intelligence. Technological entanglement (report).

Kebande, V.R., Venter, H.S., 2018. Novel digital forensic readiness technique in the cloud environment. Australian J. Forensic Sci. 50 (5), 552–591.

Kelly, L., Sachan, S., Ni, L., Almaghrabi, F., Allmendinger, R., Chen, Y.-W., 2020. Explainable artificial intelligence for digital forensics: opportunities, challenges and a drug testing case study. In: Digital Forensic Science, IntechOpen, pp. 1–20.

Kim, H., Kim, I., Kim, K., 2021a. Aibft: artificial intelligence browser forensic toolkit. Forensic Sci. Int. Digit. Investig. 36, 301091.

Kim, S., Jo, W., Lee, J., Shon, T., 2021b. AI-enabled device digital forensics for smart cities. J. Supercomput., 1–16. Publisher: Springer.

Krishnan, S., et al., 2022. Sentiment and behavioral analysis in ediscovery. Ph.D. thesis. Sam Houston State University.

Leone, M., 2021. From fingers to faces: visual semiotics and digital forensics. Int. J. Semiot. Law-Rev. Int. Sémiot. Jurid. 34 (2), 579–599.

Liu, Y., Arunachalam, S., Temme, K., 2021. A rigorous and robust quantum speed-up in supervised machine learning. Nat. Phys. 17 (9), 1013–1017.

Malhotra, S., 2023. Digital forensics meets ai: a game-changer for the 4th industrial revolution. In: Artificial Intelligence and Blockchain in Digital Forensics. Publishers, River, pp. 1–20.

Marcus, G., Davis, E., 2019. Rebooting AI: Building Artificial Intelligence We Can Trust. Vintage.

McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E., 1955. A proposal for the dartmouth summer research project on artificial intelligence. AI Mag. 31 27 (4), 12 (2006).

McKemmish, R., 2008. When is digital evidence forensically sound? In: IFIP International Conference on Digital Forensics. Springer, pp. 3–15.

McKinnel, D.R., Dargahi, T., Dehghantanha, A., Choo, K.-K.R., 2019. A systematic literature review and meta-analysis on artificial intelligence in penetration testing and vulnerability assessment. Comput. Electr. Eng. 75, 175–188.

Meissner, C.A., Kassin, S.M., 2002. "He's guilty!": investigator bias in judgments of truth and deception. Law Hum. Behav. 26 (5), 469–480.

Meissner, G., 2020. Artificial intelligence: consciousness and conscience. AI Soc. 35 (1), 225–235.

Mitchell, F., 2010. The use of artificial intelligence in digital forensics: an introduction. Digit. Evid. Electron. Signat. Law Rev. 7, 35.

Nguyen, T.N., Choo, R., 2021. Human-in-the-loop xai-enabled vulnerability detection, investigation, and mitigation. In: 2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, pp. 1210–1212.

Nissan, E., 2012. Flints, a tool for police investigation and intelligence analysis: a project by Richard Leary explained by its author. In: Computer Applications for Handling Legal Evidence, Police Investigation and Case Argumentation. Springer, pp. 767–839.

Nordby, S.K., Bjerke, A.H., Mifsud, L., 2022. Primary mathematics teachers' understanding of computational thinking. Künstl. Intell. 36 (1), 35–46.

Nowroozi, E., Dehghantanha, A., Parizi, R.M., Choo, K.-K.R., 2021. A survey of machine learning techniques in adversarial image forensics. Comput. Secur. 100, 102092.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al., 2021. The prisma 2020 statement: an updated guideline for reporting systematic reviews. Int. J. Surg. 88, 105906.

Pan, L., Batten, L., 2005. Reproducibility of digital evidence in forensic investigations. In: DFRWS 2005: Proceedings of the 5th Annual Digital Forensic Research Workshop, Digital Forensics Research Workshop, pp. 1–8.
Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P., 2020. Closing the ai accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 33–44.
Raponi, S., Oligeri, G., Ali, I.M., 2022. Sound of guns: digital forensics of gun audio samples meets artificial intelligence. Multimed. Tools Appl. 81 (21), 30387–30412.
Rowlingson, R., et al., 2004. A ten step process for forensic readiness. Int. J. Digit. Evid. 2 (3), 1–28.
Rughani, P.H., 2017. Artififical intelligece based digital forensics framework. Int. J. Adv. Res. Comput. Sci. 8 (8).
Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R., 2019. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, vol. 11700. Springer Nature.
Sanchez, L., Grajeda, C., Baggili, I., Hall, C., 2019. A practitioner survey exploring the value of forensic tools, ai, filtering, & safer presentation for investigating child sexual abuse material (csam). Digit. Investig. 29, S124–S142.
Sanyasi, M., Kumar, P., 2021. Digital forensics investigation for attacks on artificial intelligence. SPAST Abstr. 1 (1).
Schmid, S., Riebe, T., Reuter, C., 2022. Dual-use and trustworthy? A mixed methods analysis of ai diffusion between civilian and defense r&d. Sci. Eng. Ethics 28 (2), 12.
Shapiro, A., 2017. Reform predictive policing. Nature 541 (7638), 458–460.
Siau, K., Wang, W., 2018. Building trust in artificial intelligence, machine learning, and robotics. Cut. Bus. Technol. J. 31 (2), 47–53.
Sikos, L.F., 2021. AI in digital forensics: ontology engineering for cybercrime investigations. Wiley Interdiscip. Rev. Forensic Sci. 3 (3). e1394, publisher: Wiley Online Library.
Solanke, A.A., 2022. Explainable digital forensics ai: towards mitigating distrust in ai-based digital forensics analysis with interpretable models. Forensic Sci. Int. Digit. Investig.
Spranger, M., Heinke, F., Appelt, L., Puder, M., Labudde, D., 2016. MoNa: Automated identification of evidence in forensic short messages. Int. J. Adv. Secur.
Totschnig, W., 2020. Fully autonomous ai. Sci. Eng. Ethics 26 (5), 2473–2485.
Urbina, F., Lentzos, F., Invernizzi, C., Ekins, S., 2022. Dual use of artificial-intelligence-powered drug discovery. Nat. Mach. Intell. 4 (3), 189–191.
Veale, M., Binns, R., Edwards, L., 2018. Algorithms that remember: model inversion attacks and data protection law. Philos. Trans. R. Soc. A, Math. Phys. Eng. Sci. 376 (2133), 20180083.
Veldhuis, M.S., Ariëns, S., Ypma, R.J., Abeel, T., Benschop, C.C., 2022. Explainable artificial intelligence in forensics: realistic explanations for number of contributor predictions of dna profiles. Forensic Sci. Int. Genet. 56, 102632.
Zhang, W.E., Sheng, Q.Z., Alhazmi, A., Li, C., 2020. Adversarial attacks on deep-learning models in natural language processing: a survey. ACM Trans. Intell. Syst. Technol. 11 (3), 1–41.