# Neurosymbolic AI in Digital Forensics: Commonsense and Qualitative Reasoning

Alessia Donata Camarda[1]

[1]University of Calabria, Rende, 87036, Italy

**Abstract**

Technological devices have become an integral part of our daily lives. Although apparently harmless, many of these contain a huge amount of information about us which can be relevant to incriminate the culprit of a crime. Due to this reason, it is essential to include actions concerning the collection and analysis of digital evidence in the several phases of the investigative process: this is where *Digital Forensics* was born. Despite the current fame gained by the field, the context in which it is placed requires particular attention in the implementation of frameworks and methods aligned with principles such as transparency, accountability, and fairness. My research proposal aims to leverage new *Neurosymbolic Artificial Intelligence* approaches to build tools and explore the possibility of automating tasks in Digital Forensics. Traditional tools alone are currently not enough to provide valid and concrete help to the field: it is thus necessary to coordinate the use of newer methods that are increasingly present in the panorama of Artificial Intelligence and Automation to tackle new tasks or re-explore already seen ones, but from a *Trustworthy* perspective. The main ingredients useful to accomplish this task will be Commonsense and Qualitative reasoning, Answer Set Programming, and Large Language Models.

**Keywords**

Answer Set Programming, Neurosymbolic AI, Digital Forensics, Commonsense reasoning, Qualitative Reasoning

## 1. Introduction

Traditional forensics, i.e. the set of activities carried out during the life cycle of an investigation, has been an omnipresent field directly related to the presence and activities of humans. Discovering actions, decisions, or intent of specific individuals and disambiguating identities are among its main purposes, which is why removing the human component from such processes is almost impossible. Before the advent of computers, forensics primarily dealt with physical evidence, such as fingerprints, blood, or handwritten documents. However, with the spread of technology, criminal activity also changed, increasingly involving the use of computers, thus making traditional forensic techniques insufficient. To address these new forms of evidence (e.g., files, logs, and so on), the investigation process has become digital, leading to the rise of *Digital Forensics* (DF) [1, 2] . DF involves the identification, collection and analysis of digital evidence, i.e., all the information extracted and obtained using electronic instruments. The rapid development of this field has entailed the birth of a wide variety of subfields, such as *Computer Forensics*, *Network Forensics*, *IoT Forensics*, *Incident Response*, and so on. All of these deal with different aspects related to the use of computers and other technologies within criminal cases. Digital forensics is part of an even larger branch of computer science: *Cybersecurity* [3], which concerns the protection of digital systems, data, and services from malicious activities carried out by attackers. Digital forensics thus integrates knowledge and methodologies from a variety of disciplines, although it remains primarily rooted in computer science. From its birth, DF has been object of study for researchers, and new scientific discoveries have gradually driven the progress of this field. Consequently, the spread of *Artificial Intelligence* (AI) has also significantly influenced the development of tools supporting the investigative process [4, 5, 6]. Clear examples of how AI has become integral to this field include tools for analyzing electronic devices and biometric identification systems that combine facial recognition, fingerprints, iris scans, and more. Despite their usefulness, such systems are often the focus of strong criticism, especially with regard to the privacy and security of the individuals

*ICLP DC 2025: 21$^{st}$ Doctoral Consortium on Logic Programming, September 2025, Rende, Italy.*

alessiadonata.camarda@unical.it (A. D. Camarda)

involved [7]. Consequently, the call for *Trustworthy AI* (TAI), which emphasizes privacy and ethics, has gained prominence in response to these concerns. TAI aims to develop AI systems whose core values are ethics, safety, transparency, and human rights and values. Several institutional entities have proposed regulations and guidelines to be followed by AI systems. For example, the European Commission [8] has established 7 key requirements that Trustworthy AI is expected to meet. Despite these guidelines, it is not always possible to fully adhere to them. For example, deep learning-based systems cannot easily ensure transparency in their operations. Moreover, several cases have demonstrated that an improper use of such models can even result in guidelines violation [9]. This demonstrates that there are still major steps to be taken to develop systems that can fully comply with the guidelines while still remaining capable of performing complex tasks.

## 2. Related work

Several efforts have been made trying to provide meaningful contributions to the field. One of the key principles on which research is focusing is explainability. In [10], a first formal definition of Explainable Artificial Intelligence in Cyber Security is proposed. The authors also offer a detailed study about which system properties are necessary to ensure AI-powered tools aligned with legal and ethical standards. Thanks to the spread of deep learning, many tasks that until now were only performed by humans have started to be (even if partially) solved. This is also true for digital forensics and cybersecurity, which has seen a great contribution from this community. For example, in [11] it is introduced ForensicLLM, a Large Language Model fine-tuned on Question and Answering samples extracted from digital forensic literature. This model is designed to support experts in their daily work by helping them analyze forensic artifacts, answer technical questions, or suggest investigative directions. In [4], an holistic and generalized framework for DF is proposed for standardization. Many of these attempts try to integrate explainability techniques within the deep learning pipeline [12, 13]. Recent advances have also leveraged Neuro-Symbolic approaches to enhance explainability. For instance, [14] proposes a method to extract logic-based global explanations from convolutional neural networks, supporting interpretability in tasks like image classification. Efforts have not only focused on Deep learning and Machine learning, but some approaches that exploit logical formalisms have also been proposed. In this regard, we mention DigForASP [15], a COST Action that aims to create a cooperation network for exploring the potential of the application of logic-based Artificial Intelligence in the Digital Forensics field. This cooperation has produced several results [16, 17, 18]. Similarly, in [19, 20], attempts are made to formalize knowledge related to criminal investigations. These works aim to develop a structured vocabulary capable of describing investigations activities, events and digital artifacts at different levels of detail and according to the task to be performed.

## 3. Goal of the research

The main goal of my research is to leverage *Neurosymbolic artificial intelligence* [21] to bridge the current gaps in digital forensics. Since traditional tools alone are currently insufficient to provide effective and concrete support to the field, it remains necessary to exploit deep learning and, in particular, *Large Language Models* [22] to perform support tasks. On the other hand, it is necessary to adhere as closely as possible to the Guidelines for a Trustworthy AI. In this respect, logical formalisms, such as *Answer Set Programming* (ASP) [23], allow to enforce transparency and explainability while simultaneously enabling reasoning over incomplete or conflicting evidence. To address two of the most critical aspects of digital forensics - namely, the world knowledge possessed by humans and the partial, uncertain information available during crime investigations - we propose integrating commonsense and qualitative reasoning into our pipelines:

**Commonsense reasoning.** Commonsense understanding about daily life, typical movement patterns for normal or abnormal behaviors, and specific expertise related to investigative actions can be useful

for identifying contradictions between depositions and the evidence available in a case. Therefore, such knowledge must be included among the information relevant for resolving the case. Affordance knowledge, i.e., information about what can or cannot be done with an object, can be mixed with case-specific knowledge. This can help in understanding, e.g., how certain objects are linked to the actions performed by the agents involved. Taking this type of information into account is important because humans acquire such knowledge over time and take it for granted, whereas computer systems do not possess it by default. It is necessary to find ways to enable them to *learn* it. On this matter, Answer Set Programming is particularly well-suited to represent new knowledge and exceptions, enabling the construction of a declarative knowledge base that encodes commonsense and affordance reasoning and supports effective inference over it.

**Qualitative reasoning.** The data available in a given case are not always complete or quantitatively precise. Indeed, law enforcement and investigators often have to work with incomplete and insufficient data. Moreover, individuals with expertise in computer science and related fields may attempt to tamper with digital evidence to hide the truth, further compromising the reliability of the available data. To implement reasoning and analysis processes, it is therefore necessary to leverage Knowledge Representation and Reasoning tools such as qualitative reasoning, which allows reasoning about qualitative notions, e.g., *near*, *hotter than*. Qualitative reasoning can be combined with rule-based systems or relational formalisms (e.g., temporal or spatial logic frameworks) to interpret incomplete data, recognize patterns, reconstruct scenarios, perform behavioral analysis, and automate the exploration of investigative hypotheses. On this matter, Answer Set Programming is particularly well-suited, as it enables the representation of incomplete or uncertain knowledge and supports inference over symbolic relations, even in the absence of reliable or quantitative data.

## 4. Current status of the research and preliminary results

Ongoing research is addressing the aforementioned topics. The following section presents some preliminary considerations and findings obtained so far.

**Contradictions management.** Contradiction management has been object of study in many fields under a different terminology. Since forensics field deals with natural language, and therefore with its complexities, e.g. ambiguity, traditional tools cannot provide useful results. Several approaches have attempted to use Deep Learning to detect contradictions [24, 25], even in the legal field [26]. Although the results are good, they are still far from optimal. We started to study the application of Neuro-Symbolic AI to contradiction management. In this context, we propose a pipeline that relegate the role of Large Language Models to support tasks, such as commonsense knowledge extraction and translation of input sentences into structured format. Then, the reasoning phase is performed by an Answer Set Programming solver, thus allowing us to justify the gained conclusions. We managed to obtain about $84\%$ accuracy on the dataset at hand. The results are presented in the following paper [27].

**Temporal reasoning.** Time management and reasoning about events are among the fundamental issues in digital forensics. When dealing with imprecise information, the results are affected by this uncertainty, which also applies to unreliable timestamps and overlapping events, thereby significantly influencing the outcome of an investigation. Different approaches have been proposed to address temporal reasoning under uncertainty, including Allen's interval algebra [28], but these methods are often difficult to integrate into automated reasoning systems. We are currently developing a system that supports temporal reasoning by approximating the available values and considering multiple scenarios simultaneously, while accounting uncertainty and managing it. Always keeping in mind the Trustworthy AI guidelines, our goal is to propose an explainable system capable of clarifying the reasoning behind the obtained results, despite the uncertainty of the available information.

**Commonsense extraction and anomaly detection.** Neuro-Symbolic AI is a discipline that can provide valuable support in various fields and has become the focus of numerous studied. For example, it can be particularly useful in areas such as anomaly detection and commonsense reasoning. In particular, we are investigating how to detect anomalies that contradict commonsense or default assumptions, rather than identifying outliers in numerical data. The pipeline under consideration leverages LLMs to extract commonsense knowledge about objects, people, and their typical attributes. Similar work has been carried out with the aim of populating ontologies using LLMs [29]. In our proposed architecture, the reasoning phase is handled by an Answer Set Programming solver, where a set of rules detects unexpected simple objects configurations or anomalous action executions.

**Decision Support Framework for Trustworthy AI.** Digital forensics operates across multiple contexts, each posing unique challenges, which makes the demand for trustworthy AI one of the most pressing. In this context, we propose the *S*ocio-*Te*chnical framework for *T*rustworthy *A*rtificial *I*ntelligence in Digital Forensics, STeForTAI, a theoretical and methodological framework. Its conception and design emerged from several meeting of the DigForASP project. STeForTAI is grounded in Socio-Technical Systems Theory, which support the analysis of complex systems involving both technical and social components, and aligns with the European Commission's Ethics Guidelines for Trustworthy AI. The framework was formally introduced in [30] and is currently undergoing peer review.

## 5. Open issues and expected achievements

Given its inherently human-centered nature, some of the open issues in digital forensics relate to its interaction with the people involved in the use of forensic tools. In addition, the increasingly involvement of artificial intelligence in the development of such tools brings with it-self several technological and methodological challenges. Below, we outline some of the most important open issues in this field and suggest some possible contributions.

**Privacy concerns and Accountability.** Criminal cases involve actions performed by individuals; thus, all the evidence collected relates to what these people did and where they were at specific points in time. Although this information is necessary for conducting an investigation, it raises privacy concerns regarding what information is collected, how it is used and who can access it. Furthermore, much of this knowledge, even if accessed during an investigation, cannot be retained by individuals who lack authorization. However, such data would be valuable for training models or developing tools to analyze it. This results in a significant lack of datasets for researchers, who can therefore seldom test their tools on real data. In addition, this prevents the standardization of automation and analysis processes, thus hindering the reproducibility and validation of proposed methods. In this regard, this research is particularly focused on building a forensic commonsense knowledge dataset based on already existing commonsense datasets. Furthermore, anonymization techniques could enable the creation of anonymized datasets that would support the initiation of a benchmark standardization process.

**The complex task of transforming available knowledge into an useful format.** The evidence collected may include data extracted from private digital devices, surveillance cameras, or physical elements found at the crime scene. To represent complex systems, such as the environments in which we live and interact, it is necessary to create simplified representations that can be more easily managed. However, determining the appropriate level of abstraction is not always easy: a higher level of abstraction can lead to the loss of important information, whereas a lower level can result in retaining details that complicate rather than clarify the problem. In this regard, various information extraction techniques, such as entity and relations extraction, could be explored with the support of LLMs. The resulting knowledge can then be represented using Answer Set Programming.

**Lack of explainability.** Answers provided by deep learning models are often not explainable, or the explanations produced are difficult for non-experts to understand. This raises the issue of how to explain to non-experts what was done and the process that led to specific results. Explainability is essential in fields such as forensics, where both the results and the processes used to obtain them must be clear, especially to those involved (e.g., judges who must issue sentences or individuals whose lives depend on such decisions). In this respect, since this research delegates the reasoning component to explainable solvers whose results are fully deterministic and reproducible, it could explore the application of neurosymbolic techniques both for unsolved tasks and for solved tasks whose existing methods lack explainability.

## Acknowledgments

This work was partially supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

## Declaration on Generative AI

The author used GPT-4o to do grammar and spelling check and rewriting of specific sentences. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

## References

[1] M. Pollitt, A history of digital forensics, in: IFIP Int. Conf. Digital Forensics, volume 337 of *IFIP Advances in Information and Communication Technology*, Springer, 2010, pp. 3–15.

[2] R. Jones, Digital evidence and computer crime: Forensic science, computers and the internet, Int. J. Law Inf. Technol. 11 (2003) 98–100.

[3] S. Alam, Cybersecurity: Past, present and future, 2024. URL: https://arxiv.org/abs/2207.01227. arXiv:2207.01227.

[4] Z. Khalid, F. Iqbal, B. C. M. Fung, Towards a unified xai-based framework for digital forensic investigations, Digit. Investig. 50 (2024) 301806.

[5] A. Wickramasekara, F. Breitinger, M. Scanlon, Exploring the potential of large language models for improving digital forensic investigation efficiency, Forensic Sci. Int. Digit. Investig. 52 (2025) 301859.

[6] A. A. Solanke, M. A. Biasiotti, Digital forensics AI: evaluating, standardizing and optimizing digital evidence mining techniques, Künstliche Intell. 36 (2022) 143–161.

[7] W. K. Jung, H. Y. Kwon, Privacy and data protection regulations for AI using publicly available data: Clearview AI case, in: ICEGOV, ACM, 2024, pp. 48–55.

[8] European Commission, Ethics guidelines for trustworthy AI, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, 2019. [Online].

[9] J. Angwin, J. Larson, S. Mattu, L. Kirchner, Machine bias: There's software used across the country to predict future criminals. and it's biased against blacks, ProPublica (2016). [Online].

[10] S. Alam, Z. Altiparmak, Xai-cf – examining the role of explainable artificial intelligence in cyber forensics, 2024. URL: https://arxiv.org/abs/2402.02452. arXiv:2402.02452.

[11] B. Sharma, J. Ghawaly, K. McCleary, A. M. Webb, I. Baggili, Forensicllm: A local large language model for digital forensics, Digit. Investig. 52 (2025) 301872.

[12] M. Zolanvari, Z. Yang, K. M. Khan, R. Jain, N. Meskin, TRUST XAI: model-agnostic explanations for AI with a case study on iiot security, IEEE Internet Things J. 10 (2023) 2967–2978.

[13] M. Wang, K. Zheng, Y. Yang, X. Wang, An explainable machine learning framework for intrusion detection systems, IEEE Access 8 (2020) 73127–73141.

[14] P. Padalkar, H. Wang, G. Gupta, Nesyfold: A framework for interpretable image classification, in: AAAI, AAAI Press, 2024, pp. 4378–4387.
[15] S. Costantini, F. A. Lisi, R. Olivieri, Digforasp: A european cooperation network for logic-based AI in digital forensics, in: CILC, volume 2396 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 138–146.
[16] F. A. Lisi, G. Sterlicchio, Mining sequences in phone recordings with answer set programming, in: HYDRA/RCRA@LPNMR, volume 3281 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 34–50.
[17] F. A. Lisi, G. Sterlicchio, Declarative AI and digital forensics: Activities and results within the digforasp project, in: Ital-IA, volume 3486 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 437–442.
[18] J. Medina-Moreno, Digital forensics: Evidence analysis via intelligent systems and practices digforasp - CA17124. challenges and achievements: Plenary talk, in: SISY, IEEE, 2022, pp. 17–18.
[19] S. Costantini, G. D. Gasperis, R. Olivieri, Digital forensics and investigations meet artificial intelligence, Ann. Math. Artif. Intell. 86 (2019) 193–229.
[20] D. Kahvedzic, M. T. Kechadi, DIALOG: A framework for modeling, analysis and reuse of digital forensic knowledge, Digit. Investig. 6 (2009) S23–S33.
[21] M. K. Sarker, L. Zhou, A. Eberhart, P. Hitzler, Neuro-symbolic artificial intelligence: Current trends, CoRR abs/2105.05330 (2021).
[22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: NIPS, 2017, pp. 5998–6008.
[23] V. Lifschitz, Answer Set Programming, Springer, 2019.
[24] V. Lingam, S. Bhuria, M. Nair, D. Gurpreetsingh, A. Goyal, A. Sureka, Deep learning for conflicting statements detection in text, PeerJ Prepr. 6 (2018) e26589.
[25] M. de Marneffe, A. N. Rafferty, C. D. Manning, Finding Contradictions in text, in: ACL, The Association for Computer Linguistics, 2008, pp. 1039–1047.
[26] S. Surana, S. Dembla, P. Bihani, Identifying Contradictions in the Legal Proceedings Using Natural Language Models, SN Comput. Sci. 3 (2022) 187.
[27] A. D. Camarda, G. Ianni, A study on contradiction detection using a neuro-symbolic approach, in: CILC, volume 4003 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025.
[28] M. Grüninger, Z. Li, The time ontology of allen's interval algebra, in: TIME, volume 90 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017, pp. 16:1–16:16.
[29] G. Ciatto, A. Agiollo, M. Magnini, A. Omicini, Large language models as oracles for instantiating ontologies with domain-specific knowledge, Knowl. Based Syst. 310 (2025) 112940.
[30] A. Brännström, A. D. Camarda, S. Costantini, P. Dell'Acqua, C. Gallese, G. Ianni, F. A. Lisi, V. Mascardi, J. C. Nieves, Supporting trustworthiness in socio-technical frameworks with logic programming, 2024. Submitted.