Neurosymbolic AI in Digital Forensics: Commonsense and Qualitative Reasoning

Alessia Donata Camarda [a,*]

[a]University of Calabria, Rende, Italy

Abstract. Digital devices contain a huge amount of information about us, e.g., what we like to do or when we do particular actions, which can be relevant to incriminate the culprit of a crime. Indeed, they may contain undeclared information or evidence that conflicts with what has been declared. Due to this reason, the identification, analysis, and management of digital evidence must become baseline actions in the investigative process: this is where *Digital Forensics* was born. Given its intrinsic human-related nature, Digital Forensics 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 develop 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.

1 Introduction

With the spread of technology, criminal activity has changed, increasingly involving the use of computers and thus making traditional forensics techniques insufficient, i.e., the set of tools and methods used to carry out activities during the life cycle of an investigation. A new form of evidence has becomes the focus of investigations: digital evidence, e.g., files, logs, emails, which allows investigators to disambiguate identities and discover actions, decisions, or intent of specific individuals. To address these new forms of evidence, the investigation process has become digital, leading to the rise of *Digital Forensics* (DF) [9, 14]. DF involves the identification, collection, and analysis of digital evidence, i.e., all the information extracted and obtained using electronic instruments. Digital Forensics is part of an even larger branch of computer science: *Cybersecurity* [1], 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. The spread of *Artificial Intelligence* (AI) has especially influenced the implementation of tools useful for the investigative process [11, 16, 19], such as systems that analyze electronic devices or build biometric identification systems combining 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 involved [10]. Indeed, the increasing involvement of Artificial Intelligence in the development of such tools raises several problems related to the technologies used.

Privacy concerns and Accountability The management of information necessary for conducting an investigation, e.g., what involved people did or where they were at a specific moment in time, entails some privacy issues due to what data is collected, how it is used, and who can access it. Furthermore, this knowledge cannot be collected and reused by researchers who would like to train models or build tools to analyze such data. This prevents the creation of datasets essential for standardizing, reproducing and validating proposed methods.

Abstraction of the available knowledge Evidence collected could concern data extracted from private digital devices, surveillance cameras or items found directly at the crime scene. Dealing with such complex data requires the creation of simplified representations that can be more easily managed. However, understanding the right level of abstraction to use is not always easy: a higher level of abstraction can lead to the loss of important information, while a lower one can lead to retaining details that complicates rather than helping to solve the problem.

Lack of explainability Legal users and non-experts usually require a narrative reconstruction of the sequence of the events and the meaning of digital evidence, in order to understand the impact of discovered knowledge on the resolution of the case. This poses the problem of how to explain to non-experts what was done and the process that led to specific outcomes. Explainability is then a necessity in fields such as forensics, where the results obtained and the processes used to obtain them must be clear especially to the people involved (e.g., judges who must issue sentences or individuals whose lives depend on such decisions).

In this respect, *Trustworthy AI* (TAI) has gained prominence in response to these concerns. TAI aims to develop AI systems whose core values include ethics, safety, transparency, and human rights. Several institutional entities have proposed regulations and guidelines to be followed by AI systems. For example, the European Commission [7] has established seven 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

[*] Email: alessiadonata.camarda@unical.it

operations. Moreover, several cases have demonstrated that the improper use of such models can even results in guidelines violation [2]. This demonstrates that there are still major steps to be taken to develop systems that can fully comply with the guidelines while remaining capable of performing complex tasks.

2 Goal of the research

My research proposal aims to reduce the gaps currently present in Digital Forensics leveraging *Neurosymbolic Artificial Intelligence* [15]. Specifically, on the one hand, I propose using Deep Learning and, in particular, *Large Language Models* [18] to perform support tasks, while on the other hand, logical formalisms, such as *Answer Set Programming* [12] can be used to enforce transparency and explainability, and enable reasoning over incomplete or conflicting evidence. Since traditional tools alone are currently insufficient to provide effective and concrete support to the field, it is still necessary to leverage Deep Learning but, at the same time, it is necessary to adhere to the Guidelines for a Trustworthy AI as much as possible. 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 Humans think and act based on commonsense understanding of daily life and typicalities, i.e., what is considered typical or not. This kind of knowledge can be useful to identify contradictions between depositions and the evidence available in a case; therefore, it should be included among the information relevant for resolving the case. Additionally, affordance knowledge, i.e., information about what can or cannot be done with an object, can be integrated with case-specific knowledge. Taking this type of information into account is essential because humans acquire it over time and take it for granted, whereas computer systems do not possess it by default. It is therefore necessary to find ways to enable them to *learn* it.

Qualitative reasoning Police and investigators often deal with incomplete or quantitatively imprecise data. Furthermore, 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. Reasoning on qualitative notions, such as *near*, *hotter than*, *faster*, is then necessary to analyze and draw conclusions from the evidence. Qualitative reasoning, a subfield of Knowledge Representation and Reasoning, can be combined with rule-based systems or relational formalisms (e.g.,, temporal or spatial logic frameworks) to automate the exploration of investigative hypotheses.

3 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, often under a different terminology. Since forensics field deals with natural language and its inherent complexities, such as ambiguity, traditional tools cannot always yield reliable results. Several approaches have attempted to use deep learning to detect contradictions [6, 13], even in the legal field [17]. Although the results are promising, they still far short of optimal performances. We started to study the application of NeuroSymbolic AI to contradiction management. In this context, we propose a pipeline that relegates the role of Large Language Models to support tasks, such as commonsense knowledge extraction and translating input sentences into structured format. The reasoning phase is then carried out by an Answer Set Programming solver, thus enabling the justification of the derived conclusions. We achieved an accuracy of approximately 84% on the dataset at hand. The results are presented in the following paper [4].

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 [8], 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 NeuroSymbolic AI is a discipline that can provide valuable support in various fields and has become the focus of numerous studies. For example, it can be 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 [5]. 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 [3].

Acknowledgements

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

References

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

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

[3] A. Brännström, A. D. Camarda, S. Costantini, P. Dell'Acqua, C. Gallese, G. Ianni, F. A. Lisi, V. Mascardi, and J. C. Nieves. Supporting trustworthiness in socio-technical frameworks with logic programming. Submitted, 2024.

[4] A. D. Camarda and G. Ianni. A study on contradiction detection using a neuro-symbolic approach. In *CILC*, volume 4003 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2025.

[5] G. Ciatto, A. Agiollo, M. Magnini, and A. Omicini. Large language models as oracles for instantiating ontologies with domain-specific knowledge. *Knowl. Based Syst.*, 310:112940, 2025.

[6] M. de Marneffe, A. N. Rafferty, and C. D. Manning. Finding Contradictions in text. In *ACL*, pages 1039–1047. The Association for Computer Linguistics, 2008.

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

[8] M. Grüninger and Z. Li. The time ontology of allen's interval algebra. In *TIME*, volume 90 of *LIPIcs*, pages 16:1–16:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.

[9] R. Jones. Digital evidence and computer crime: Forensic science, computers and the internet. *Int. J. Law Inf. Technol.*, 11(1):98–100, 2003.

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

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

[12] V. Lifschitz. *Answer Set Programming*. Springer, 2019.

[13] V. Lingam, S. Bhuria, M. Nair, D. Gurpreetsingh, A. Goyal, and A. Sureka. Deep learning for conflicting statements detection in text. *PeerJ Prepr.*, 6:e26589, 2018.

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

[15] M. K. Sarker, L. Zhou, A. Eberhart, and P. Hitzler. Neuro-symbolic artificial intelligence: Current trends. *CoRR*, abs/2105.05330, 2021.

[16] A. A. Solanke and M. A. Biasiotti. Digital forensics AI: evaluating, standardizing and optimizing digital evidence mining techniques. *Künstliche Intell.*, 36(2):143–161, 2022.

[17] S. Surana, S. Dembla, and P. Bihani. Identifying Contradictions in the Legal Proceedings Using Natural Language Models. *SN Comput. Sci.*, 3(3):187, 2022.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *NIPS*, pages 5998–6008, 2017.

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