An ML-BDI Reasoner to Support Crime Investigation in Digital Forensics

Extended Abstract

Guilherme Dall'Agnol Deconto
PUCRS
Porto Alegre, Brazil
g.dallagnol@edu.pucrs.br

Leonardo dos Santos Teixeira
IFRS
Porto Alegre, Brazil
2021004121@aluno.restinga.ifrs.edu.br

Roben Castagna Lunardi
IFRS and PUCRS
Porto Alegre, Brazil
roben.lunardi@zonanorte.ifrs.edu.br

Rafael C. Cardoso
University of Aberdeen
Aberdeen, United Kingdom
rafael.cardoso@abdn.ac.uk

Felipe Meneguzzi
University of Aberdeen and PUCRS
Aberdeen, United Kingdom
felipe.meneguzzi@abdn.ac.uk

Avelino Francisco Zorzo
PUCRS
Porto Alegre, Brazil
avelino.zorzo@pucrs.br

ABSTRACT

In Digital Forensics, investigators may rely on IoT traces from smart environments to reconstruct whether and how many people were present in a room, yet sensor outages and noise can produce misleading inferences. We introduce an ML-BDI reasoner that combines a Random Forest occupancy estimator with a BDI agent that selects symbolic fallback plans during sensor failures and applies confidence-aware policies, including abstention in low-reliability conditions. The system records an auditable trace of each decision (plan id, timestamp, RF confidence, sensor state) to support forensic review. Overall, the reasoner accepts high-confidence predictions and otherwise falls back to symbolic plans or abstains.

KEYWORDS

Random Forest, BDI Agent, Digital Forensics, IoT

ACM Reference Format:
Guilherme Dall'Agnol Deconto, Leonardo dos Santos Teixeira, Roben Castagna Lunardi, Rafael C. Cardoso, Felipe Meneguzzi, and Avelino Francisco Zorzo. 2026. An ML-BDI Reasoner to Support Crime Investigation in Digital Forensics: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/ZGRS2204

1 INTRODUCTION

Digital forensics has been facing a practical shift: the volume and variety of potential evidence continue to grow [14]. Investigators now often examine not only traditional storage media but also mobile devices, cloud services, and network traces. This creates a problem in the everyday sense: too much data, too many formats, too little time, yet the need to adhere to forensic principles [14].

Smart buildings generate continuous streams from many devices, which can be useful as evidence, but these environments are also fragile. Devices fail, networks fail, firmware changes, and sometimes disruptions are deliberate. In IoT attacks, anti-forensic actions (e.g., disabling devices) can compromise both the availability and the integrity of data. For an investigator, that means conclusions may depend on partial, noisy, or inconsistent traces [1, 5, 22].

At the same time, there is a push to use Machine Learning (ML) in digital forensics because it helps with tasks that do not scale manually, such as triage, pattern recognition, anomaly detection, and extracting information from messy data [4, 15, 20]. But in forensic work, speed is not enough. The path from raw evidence to a conclusion must remain auditable [12]. In practice, that means the pipeline matters. Ingestion, training, validation, and post-hoc analysis need to be organised in a way that supports later checking and reporting [3, 8, 17, 18]. This is one reason why interpretability and hybrid designs feature in forensic discussions: it is not only about accuracy; it is also about the ability to justify decisions.

In this context, we propose an ML-BDI reasoner for smart-building investigations that combines a Random Forest (RF) model for ML-based perception with symbolic decision policies to support inference under sensor unreliability while maintaining an auditable record of how the agent makes decisions.

We choose a Belief-Desire-Intention (BDI) agent as the symbolic layer because it provides a way to encode and execute explicit policies that are easy to inspect and report [7, 10]. Our implementation uses the Jason platform [6] as the symbolic component, which implements BDI agents on top of the AgentSpeak(L) programming language [19]. In our setting, the agent's beliefs represent the current sensor state and the model's confidence, and its plans implement conservative actions, such as falling back to sensor-only rules or abstaining when reliability is low. This meets the needs of digital forensics, provides explainable conclusions, and aligns with our goal of producing an auditable record of how the agent reaches each decision.

2 BACKGROUND & RELATED WORK

Digital forensics is concerned with collecting, analysing, and preserving digital data for use as evidence. It is also a field where reliability and admissibility are constantly in the background: if a method is not reproducible, or if its outputs cannot be explained and documented, it is harder to trust and harder to defend. This is why classic forensic process models emphasise steps such as acquisition, examination, analysis, and reporting [1, 8, 18].

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/ZGRS2204

Digital forensics employs AI methods to handle large, heterogeneous datasets. Neural and other data-driven approaches have demonstrated strong empirical performance across several forensic tasks. Still, two issues remain in forensic settings: explainability and generalisation. A model can perform well on a benchmark and still be hard to audit, and in legal contexts, the "why" can matter as much as the "what" [4, 11, 15, 20].

Symbolic approaches are usually easier to inspect and align naturally with the way investigators describe reasoning, (if x and y then consider z). This fits well with forensic reporting. However, symbolic systems can encounter scalability and coverage issues as the number of situations increases, particularly when plans and rules must be created and maintained manually [9, 13].

This tension motivates hybrid ML/symbolic systems, which use learning for perception and pattern extraction, while using symbolic reasoning to enforce explicit decision policies and provide clearer traces for review. Our work follows this direction by combining an ML estimator for occupancy with a BDI agent that can fall back to symbolic plans when sensor conditions reduce the reliability of the ML output. The agent also keeps a record of the decision path for later forensic inspection [12, 17, 21].

3 ML-BDI REASONER

We split the dataset into training (80%) and test (20%). The Random Forest model is trained on the training partition, while the test partition is progressively degraded to simulate sensor unavailability. The degradation step applied Missing Completely at Random (MCAR) per-cell random erasure and time-windowed sensor outages to simulate realistic sensor dropouts [16]. During inference, the RF processes each timestamped instance to obtain an occupancy prediction and a confidence score, which are then sent to a Jason BDI agent together with the degraded sensor readings. When confidence is high, the agent accepts the model predictions, and when it is low, it triggers symbolic inference plans to produce a symbolic estimate. Figure 1 summarises the ML-BDI pipeline.

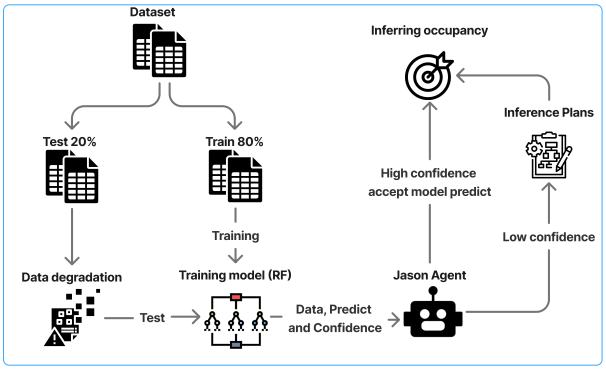

Figure 1: ML-BDI reasoner architecture.

The ML-BDI reasoner defines how it combines predictions and symbolic inference under degraded conditions. To evaluate the practical impact of this decision process, we present preliminary results obtained under progressive sensor degradation, focusing on prediction accuracy, the behaviour of the intervention, and the overall reliability of the system.

Table 1 presents the summary per seed of the BDI agent, showing how often it intervened, how many corrections were successful, and how these results compare with RF errors. The number of interactions between seeds remains fairly consistent, and the BDI regularly corrects a large share of RF misclassifications. Although the exact counts vary slightly between seeds, the overall result is clear: the BDI contributes meaningful corrections precisely when the RF struggles, without over-activating or introducing unnecessary decisions. These results demonstrate that the hybrid reasoner performs consistently across different random configurations, maintaining a stable and balanced contribution to overall accuracy.

On average, the agent intervened in about 210 cases per run, resolving roughly 75% of them. In most seeds, the number of BDI corrections even exceeded the number of RF errors (e.g., seeds 42, 101112, and 192021), showing that symbolic reasoning not only compensates for RF misclassifications but also promotes more consistent decision patterns overall. The small variation observed across seeds further indicates that the hybrid reasoner remains stable under different random initialisations and degradation scenarios, demonstrating both reliability and robustness.

The difference between the number of BDI corrections and RF errors (*BDICorrect* − *RFError*) shows a consistently positive margin across all runs, ranging from +35 to +101. This pattern confirms that the symbolic layer provides genuine corrective value rather than merely mirroring the ML model's predictions. Even in seeds where the RF had already performed well (e.g., 222324), the BDI maintained a measured level of activation, intervening only when necessary. This behaviour demonstrates that the hybrid architecture effectively adapts to varying model conditions, enhancing overall reliability without introducing redundant decisions.

Table 1: Summary per seed of the BDI agent's performance.

Seed	BDI Interactions	BDI Correct	BDI Error	RF Error	BDI Correct - RF Error
42	227.0	199.0	28.0	133.0	101.0
123	202.0	149.0	53.0	95.0	57.0
456	233.0	159.0	74.0	133.0	89.0
789	240.0	182.0	58.0	96.0	57.0
101112	236.0	187.0	49.0	115.0	90.0
131415	223.0	163.0	60.0	87.0	62.0
161718	206.0	152.0	54.0	76.0	48.0
192021	228.0	171.0	57.0	123.0	89.0
222324	143.0	108.0	35.0	50.0	35.0
252627	218.0	166.0	52.0	97.0	65.0

4 FINAL CONSIDERATIONS

This work developed an ML-BDI reasoner designed to support occupancy inference in digital forensic scenarios involving IoT sensors. From a digital forensic perspective, the approach improves the reliability of evidence. This is particularly relevant in IoT investigations, where sensor availability and integrity cannot be guaranteed.

As future work, we plan to evaluate the approach on additional datasets and sensing configurations, and to extend the architecture toward neuro-symbolic reasoning [2] and broader digital forensic scenarios.

ACKNOWLEDGMENTS

This study was partially funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – Brazil, and by the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) through grant no. 22/2551-0000841-0 (INOVA-RS). This project was also funded by CNPq/MCTI/FNDCT N° 22/2024, project number 444727/2024-8. In addition, Roben Lunardi is supported by IFRS and holds a postdoctoral fellowship from CAPES (PIPD/CAPES). Avelino F. Zorzo receives a grant from CNPq/Brazil - number 308752/2025-2. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior – Brasil (CAPES) – Finance Code 001.

REFERENCES

[1] Abdulghani Ali Ahmed, Khalid Farhan, Waheb A. Jabbar, Abdulaleem Al-Othmani, and Abdullahi Gara Abdulrahman. 2024. IoT Forensics: Current Perspectives and Future Directions. *Sensors* 24, 16 (2024). https://doi.org/10.3390/s24165210

[2] Hilal Al Shukairi and Rafael C. Cardoso. 2023. ML-MAS: A Hybrid AI Framework for Self-Driving Vehicles. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems* (London, United Kingdom) *(AAMAS '23)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1191–1199. https://doi.org/10.5555/3545946.3598762

[3] Humaira Arshad, Aman Bin Jantan, and Oludare Isaac Abiodun. 2018. Digital forensics: Review of issues in scientific validation of digital evidence. *Journal of Information Processing Systems* 14, 2 (2018). https://jips-k.org/digital-library/2018/14/2/346

[4] Konstantia Barmpatsalou, Tiago Cruz, Edmundo Monteiro, and Paulo Simoes. 2018. Mobile Forensic Data Analysis: Suspicious Pattern Detection in Mobile Evidence. *IEEE Access* 6 (2018), 59705–59727. https://doi.org/10.1109/ACCESS.2018.2875068

[5] Johnny Bengtsson. 2025. The ghost in the building: Non-invasive spoofing and covert attacks on automated buildings. *Forensic Science International: Digital Investigation* 52 (2025), 301880. https://doi.org/10.1016/j.fsidi.2025.301880

[6] Rafael H Bordini, Jomi Fred Hübner, and Michael Wooldridge. 2007. *Programming multi-agent systems in AgentSpeak using Jason*. Vol. 15. John Wiley & Sons.

[7] Rafael C. Cardoso and Angelo Ferrando. 2021. A Review of Agent-Based Programming for Multi-Agent Systems. *Computers* 10, 2 (2021), 16. https://doi.org/10.3390/computers10020016

[8] Luca Caviglione, Steffen Wendzel, and Wojciech Mazurczyk. 2017. The Future of Digital Forensics: Challenges and the Road Ahead. *IEEE Security & Privacy* 15, 6 (November 2017), 12–17. https://doi.org/10.1109/MSP.2017.4251117

[9] Stefania Costantini, Giovanni De Gasperis, and Raffaele Olivieri. 2019. Digital forensics and investigations meet artificial intelligence. *Annals of Mathematics and Artificial Intelligence* 86, 1 (01 Jul 2019), 193–229. https://doi.org/10.1007/s10472-019-09632-y

[10] Lavindra de Silva, Felipe Meneguzzi, and Brian Logan. 2020. BDI Agent Architectures: A Survey. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 4914–4921. https://doi.org/10.24963/ijcai.2020/684

[11] Guilherme Dall'Agnol Deconto, Avelino Francisco Zorzo, Daniel Bertoglio Dalalana, Edson Oliveira, and Roben Castagna Lunardi. 2024. Machine Learning for Forensic Occupancy Detection in IoT Environments. In *Good Practices and New Perspectives in Information Systems and Technologies*, Álvaro Rocha, Hojjat Adeli, Gintautas Dzemyda, Fernando Moreira, and Aneta Poniszewska-Marańda (Eds.). Springer Nature Switzerland, Cham, 102–114. https://doi.org/10.1007/978-3-031-60215-3_11

[12] Brandon L. Garrett and Cynthia Rudin. 2023. Interpretable algorithmic forensics. *Proceedings of the National Academy of Sciences* 120, 41 (2023), e2301842120. https://doi.org/10.1073/pnas.2301842120

[13] Alexander E. Grojek and Leslie F. Sikos. 2022. *Ontology-Driven Artificial Intelligence in IoT Forensics*. Springer International Publishing, Cham, 257–286. https://doi.org/10.1007/978-3-031-10706-1_12

[14] Alessandro Guarino. 2013. Digital Forensics as a Big Data Challenge. In *ISSE 2013 Securing Electronic Business Processes: Highlights of the Information Security Solutions Europe 2013 Conference*. Springer Fachmedien Wiesbaden, 197–203.

[15] Hans Henseler, Jop Hofste, and Maurice van Keulen. 2013. Digital-Forensics Based Pattern Recognition for Discovering Identities in Electronic Evidence. In *2013 European Intelligence and Security Informatics Conference*. 112–116. https://doi.org/10.1109/EISIC.2013.24

[16] Nwamaka U. Okafor and Declan T. Delaney. 2021. Missing Data Imputation on IoT Sensor Networks: Implications for on-Site Sensor Calibration. *IEEE Sensors Journal* 21, 20 (2021), 22833–22845. https://doi.org/10.1109/JSEN.2021.3105442

[17] Edson OliveiraJr, Avelino F. Zorzo, and Charles Varlei Neu. 2020. Towards a conceptual model for promoting digital forensics experiments. *Forensic Science International: Digital Investigation* 35 (2020), 301014. https://doi.org/10.1016/j.fsidi.2020.301014

[18] Sriram Raghavan. 2013. Digital forensic research: current state of the art. *CSI Transactions on ICT* 1, 1 (01 Mar 2013), 91–114. https://doi.org/10.1007/s40012-012-0008-7

[19] Anand S Rao. 1996. AgentSpeak (L): BDI agents speak out in a logical computable language. In *European workshop on modelling autonomous agents in a multi-agent world*. Springer, 42–55.

[20] Faisal Shahzad, Abdul Rehman Javed, Zunera Jalil, and Farkhund Iqbal. 2020. *Cyber Forensics with Machine Learning*. Springer US, New York, NY, 1–6. https://doi.org/10.1007/978-1-4899-7502-7_987-1

[21] Leslie F. Sikos. 2021. AI in digital forensics: Ontology engineering for cybercrime investigations. *WIREs Forensic Science* 3, 3 (2021), e1394. https://doi.org/10.1002/wfs2.1394

[22] Jean-Paul A. Yaacoub, Hassan N. Noura, Ola Salman, and Ali Chehab. 2022. Advanced digital forensics and anti-digital forensics for IoT systems: Techniques, limitations and recommendations. *Internet of Things* 19 (Aug 2022), 100544.