International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 8
Aug 2026, PP 1-10

Forensic Readiness for Autonomous AI Agents: Addressing Attribution, Accountability, and Cybercrime Investigation Requirements

[1]Dr. Pramod Kumar,
Associate Professor, Faculty of Law, SAM Global University, Raisen, Madhya Pradesh, India[1]

[2]Mr. Mohit Shrivastava,
Assistant Professor, Faculty of Law, SAM Global University, Raisen, Madhya Pradesh, India[2]

[3]Mr. Saood Iqbal khan,
Assistant Professor, Faculty of Law, SAM Global University, Raisen, Madhya Pradesh, India[3]

[4]Mr. Lakshya Bhardwaj,
Assistant Professor, Faculty of Sciences, SAM Global University, Raisen, Madhya Pradesh, India[4]

Abstract **– There are many autonomous artificial intelligence agents which can plan, carry out and adapt multi-step digital activities on their own. This is far outpacing the legal and technical tools that the law enforcement relies on to investigate and prosecute cybercriminals. Self-organizing agents make a series of coordinated decisions, call external agents, and produce emergent behaviour that is non-linear and not totally programmed or even anticipated by the programmer. This poses a challenge to forensically trace actions back to a human agent. The paper will examine the state of forensic readiness, or the organisational and technical ability to gather and present digital evidence for evidentiary purposes in the context of a digital investigation, and question its suitability for an agentic AI world. This international and national legal instrument provides a survey of the Council of Europe's Budapest Convention on Cybercrime, the EU Artificial Intelligence Act, data protection law and comparative evidence rules from the US and India. The purpose is to evaluate the current status of the immunity and liability in the case of an autonomous agent performing the proximate act instead of a natural person. A layered approach offering mandatory action logging, cryptographic provenance mechanisms, clear allocation of liability among developers, deployers and operators, as well as international cooperation mechanisms scaled to machine speed are required for the forensic readiness of autonomous agents. The paper finishes by providing a forensic readiness framework for the agentic AI lifecycle, as well as legislative, regulatory and institutional reforms to fill the upsurge of an accountability gap.**

Keywords **– forensic readiness; autonomous AI agents; digital evidence; attribution; cybercrime investigation; accountability; algorithmic logging; electronic evidence admissibility**

I. INTRODUCTION

For a long time, digital Forensics has been based on the premise that, regardless of the complexity of the tools used, any criminal or tortious act within cyberspace is a tangible result of a single human choice. An investigator investigating an intrusion, fraud or data exfiltration event has generally been able to work the chain of causation backwards from a keystroke, an executed script or a submitted command to the individual that caused it. This is violated by assuming non-autonomous intelligence agents. The high-level objective that the systems must receive, the ability to decompose this objective into sub-objects, the ability to select from and invoke tools (e.g., web browser, code interpreter, APIs, file systems), and further iterate on the results, with little or no human supervision at each intermediate step. In the earlier generations of software, instructions were performed in some way that, at least in theory, was completely defined by the software's programmer. By comparison, autonomous agents using LLMs behave probabilistically, are context-dependent, and sometimes emergent. If the exact sequence of actions required to achieve some goal was not predicted by the designer (or definer or deployer) of the system, then the behaviour is emergent.

The transition is very important in the cybercrime investigation. If the bad guy uses an autonomous agent, or an authorized party assumes control of an autonomous agent, for the purpose of reconnaissance, creating phishing material, looking for vulnerabilities, or exfiltrating information, then the impact of the attack is the same as if it was conducted by a more traditional means of a cybercrime, but the evidence trail will differ in nature. Police will have to deal with logs that are spread out among the technology makers, non-deterministic output that makes reproducing the logs difficult, and a growing number of possible first responders – model makers, agent orchestration platforms and end-stage deployers – who are all equally likely to claim responsibility for the actions of the agent. Forensic readiness is the organization and technical readiness in advance that allows an organization to promptly and effectively retrieve and retain legitimate digital evidence when needed. Good practice within the normal flow of information security governance is to be Forensic ready. In the article, it is suggested that agentic AI should be adapted to the same context, namely the developer or deployer of the system who can most

effectively prepare for forensic readiness is likely to be the best candidate for being investigated in any investigation that might arise.

The analysis goes on in a further ten parts. Part II details the technical and operational characteristics of the autonomous agents that can affect the topology of cybercrime. The third part contains the conceptual underpinnings of forensic readiness. Delegated, multi-layered agentic action creates an attribution problem that is discussed in Part IV. Part V discusses existing international and domestic legislation, pertinent to the issue of forensic readiness, at the cybercrime, artificial intelligence and data protection levels. Part VI discusses the admissibility of agents' records as evidence under comparative rules of evidence. The candidate accountability models for distributing liability among the actors that deploy an autonomous agent are discussed in Section VII. Part VIII presents a framework to ensure real readiness for the agentic lifecycle. The limitations of the framework and recommendations are presented in Parts IX and X. Part XI concludes.

II. AUTONOMOUS AI AGENTS AND THE CHANGING TOPOLOGY OF CYBERCRIME

For the purposes of this paper, a software system that has a machine learning model, which is given a natural-language or structured objective, which generates a plan to achieve the objective, which uses one or more tools or application programming interfaces to carry out part of the plan, which observes the results of the executions, and which repeats this process without human approval of intermediate steps is an autonomous agent. This architecture, also known as an agentic loop, is different from previous automation in three ways that affect forensic readiness. The logic behind the action of the system is not clearly understood, not so much by the end user as by the system's creator. That is because the model does not produce the intermediate steps in an audit-able manner, but in a stochastic manner. To make matters worse, calls are often orchestrated by agents using several third-party services. It involves a call reaching the logs of a single model application programming interface (“API”) provider, a cloud compute host, a browsing tool, and one or more downstream websites or systems, where the data retention policies, contractual obligations and legal regimes on these systems differ. Also, agents are able to make other agents and give away sub tasks. This means that an investigator has to find increasingly specific decision points to build up a causal explanation.

These attributes create further attack opportunities and lower the costs of some forms of cybercrime. Autonomous agents can be programmed, either by their rightful owner or by an attacker that has managed to prompt or manipulate the agent, to automatically reconnoiter target systems, create enormous amounts of customised phishing content and even conduct credential-stuffing and vulnerability-scanning operations at a scale and speed that would not be possible for a human adversary to achieve manually. Possibly even more important is the possibility of having an agent act in a legitimate manner being misled through various devious ways of manipulating the data or the content on which it acts so that the person operating it would not have wanted it to do that act or could not have imagined that it would do that act. The potential for differences in the intent of the actor and instrumentality makes the conventional forensic and legal rule that the actor whose intent is responsible for the actions of a system, is the one in charge of the system, a difficult one to maintain. Much of current cybercrime law is based on this presumption, including the requirement for the proof of unauthorised access or fraudulent intent.

III. FORENSIC READINESS: CONCEPTUAL FOUNDATIONS

According to the information security literature forensic readiness is the organisation's capacity to maximize the value of the use of digital evidence while minimizing the cost of investigation. The idea is not to attempt to collect digital evidence as a response to the discovery of an event, but to develop and enhance the technical and procedural capability to gather, preserve and authenticate digital evidence for use at a level that is legally admissible in advance of any event being discovered. The original formulation has ten steps: Identify business scenarios where digital evidence is required; Identify data sources; Identify the requirements for collecting digital evidence; Establish a lawful capability of collection; Implement secure data collection and storage policies; Implement monitoring that can detect incidents and incidents of digital evidence; Train employees to understand the sensitivity of digital evidence; Document cases appropriately for use in legal or disciplinary process.

I'm sorry, but your paraphrasing request is too short for me to process. The ISO/IEC 27036-2:2014 the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) -standard sets out idealised models for incident investigation processes. In the standard, the procedure of digital investigation is called an incident investigation. The models, incidents, and other vocabulary is explained in the standard. The models encompass the entire life cycle of a digital investigation ranging from pre-incident preparation to initialisation, evidence acquisition, investigation, and closure. There are two other standards. A standard that does not only relate to forensic methods is one dealing with investigation methods and techniques. The other standard is a governance framework, which gives organisations a high-level governance framework to support strategic pre-incident forensic preparation. The National Institute of Standards and Technology (NIST) has provided guidance on how incorporated forensic methods should be used in the process of evidence collection as part of an incident response investigation, including collecting evidence from all possible sources, such as logs, network traffic, and file systems, and preserving the evidential value of the evidence.

These traditional IT regulatory systems were designed for deterministic systems, which generate evidentiary data, and in which one custodian can have a significant control. Both of the assumptions that are limiting autonomous agents. A forensic readiness program, built on static log formats and single custodian data stores, does not program without modification to encode the distributed, probabilistic and multi-party nature of agentic computation. The remainder of this paper addresses the preservation of the security goals of traceability, integrity and admissibility in this changed world in the context of forensic readiness.

IV. THE ATTRIBUTION PROBLEM: TRACING CONDUCT TO ACCOUNTABLE ACTORS

There are two levels of attribution in digital forensics. Technical attribution is the process of determining which device, account, network address, system or other computer-related entity is responsible for the act. Legal attribution relates to the identification of a natural person, a corporation or another such entity capable of having legal rights and obligations; the act's actor to whom the act shall be attributed. The conventional cybercrime investigation is based on the close interaction between these two levels of ‘inference’. Where the technical trail is established then the legal attribution is based on the ownership or control of the involved account or device, with the exception of such defences as use by another party.

The new layer of Autonomous agents is neither technical nor legal. The agent itself is the actor performing the step in question, e.g. the one that submits the request to the target server, or the one that drafts the message. The agent in itself has no legal capacity and is not liable for a crime or tort. Legal attribution involves determining who is legally accountable for a particular product in the causal sequence of design, training, fine-tuning, deployment, configuration and operation. It is plausible that five different actors could be involved in a single incident; the developer of the underlying foundation model, anyone who fine-tuned or otherwise modified the model for a particular use, the developer of the agentic orchestration layer which gives the model access to tools, the deployer who integrated the agent into a live product or service, and the end user or operator who provided the objective which the agent acted upon. None of these actors will see the behaviour of an agent in its entirety, and each one can justify his/her own belief that it was not the agent's fault.

Delegation and sub-agent spawning are involved in delegating tasks to other agents. Sub-agents can each have their own context window, tools and loggers footprint. These practices have the effect of spreading the focus of the principal agent. So, for this reason, these files will often be spread across various policies and systems of various corporate custodians in various jurisdictions, with their metadata having to be correlated by the investigator, and each jurisdiction imposing on the custodian different rules regarding the retention and disclosure of the data. If there is only a temporary retention of an agent's intermediate reasoning trace, and the evidence is discarded at the end of the session, then if the investigation begins, it seems that the legal basis for attribution may just not exist. The problems of attribution are not necessarily of the nature of tried to apply pre-existing doctrines on new issues because of the evidence survival problem. This is the problem which forensic readiness is aiming to mitigate.

V. REGULATORY AND LEGAL FRAMEWORKS GOVERNING FORENSIC READINESS

The Budapest Convention - The key international treaty on cybercrime investigation which the Council of Europe has drawn up. The Convention, which came into force on 1 July 2004, has three closely interrelated goals: further harmonisation of the substantive criminal law in relation to a number of computer and content-related offences, such as offences involving unlawful access, interference with data and computer-related fraud; the creation of domestic powers to investigate such offences, including powers for the expeditious preservation of stored computer data and production orders; and the framework for international co-operation covering – in particular – a tool for countries to use in their bilateral and multilateral mutual legal assistance efforts, taking account of the cross-border nature of electronic evidence.

The Convention wasn’t framed with autonomous programs in mind and in any way makes no distinction between behavior executed by human operators and behavior with mediation by automated or semi-autonomous systems. However, the process towards expedient conservation and cross border cooperation is still the main lever through which evidence from foreign service providers, such as those who operate agentic AI platforms, would in practice have to be acquired.

The Artifical Intelligence Act, Regulation (EU) 2024/1689 is based on a risk-based regulatory framework and contains an obligation for systems that are considered high-risk including the forensic readiness aspect. High-risk AI systems should be able to automatically log events throughout the systems' lifespan. This is generally referred to as logging. The logging shall be designed so that it is possible to detect when the system may be at risk or has a significant change. Furthermore, the logging capability shall enable the possibility of monitoring the system after deployment and monitoring of the operation of the system on the deployer side. If a high-risk system is provided, the logs created by the system will always need to be under the control of the provided system. The deployment of the application is the responsibility of the deployer, who will have to manage the automatically generated logs separately and work with competent authorities. If these record keeping requirements are not met, they may be subject to a regulation scheme penalty. The Artificial Intelligence (AI) Act framework classifies high-risk applications, but the principle behind the classification – that systems that could

have material effects on an individual or infrastructure require a record of how they function that can be made easily available and updated in real time – is the most relevant existing regulatory framework for AI systems to be subject to "forensic readiness" requirements.

There are two points of contact between forensic readiness and documents regarding data protection, such as the General Data Protection Regulation. The accountability principle imposes a requirement that the controllers be able to justify compliance with data processing obligations; in reality, the requirement is similar to the one found in the forensic readiness principle and calls for the same level of contemporaneous record keeping. They are two distinct sets of laws, however. The stipulative statement, then, which forbids decisions to be taken based exclusively on automated processing without any human participation, has a double effect on the person: legal and/or equally. This provides an incentive for organisations that are deploying autonomous agents (aka AI systems without human intervention) to keep records. This can include representations of the degree of and type of human intervention that has actually taken place in a specific decision. Absence of such material makes it equally hard to rebut a claim like that of a decision being one that was taken without adequate human involvement. Data protection requirements also conflict with forensic readiness, with the doctrine of data minimisation leading to a desire to not store large volumes of information about behaviour for extended periods, which is explored further in Part IX.

18 U.S.C. section 1030 is part of the United States federal law and is called the Computer Fraud and Abuse Act. This Act makes it an offence to access the protected computer without permission. Further, it also criminalizes other behavior that is related to the same. Moreover, investigators usually obtain electronic evidence via the third-party service provider. Also, it is done under the Stored Communications Act. This Act is 18 U.S.C. section 2701 which is and the following provisions. The Computer Fraud Act and the Wiretap Act; neither statute nor anything on the record indicates an intention that courts should resolve cases where an autonomous agent physically performs the proximate act of unauthorised access and not directly by a human defendant. There has not yet been a case before which has been resolved by courts. This issue of the constitutional protection of the comprehensive behavioural records that agentic systems can produce has, however, been indirectly raised by the Supreme Court. In the Carpenter case, the Court determined that the government's historical collection of cell-site information constituted a search" for Fourth Amendment purposes. The basis for this conclusion was that the records of such place were "intimate windows into a person's life"- and therefore covered by the Fourth Amendment. A previous judgement of the court in Riley v. California held that the search of a cell phone incident to arrest requires a warrant since digital devices implicate qualitatively greater privacy interest than physical objects due to their capacity and pervasiveness. In both cases, the detailed, cumulative digital records that an agentic forensic readiness regime would generate appear to be likely to come under more constitutional challenge in the future when government seeks access. It is this consideration that any legislative logging requirement in the United States would have to consider.

Cybercrime is given the main substantive provisions in the Information Technology Act (IT Act), 2000, of India. The IT Act also provides for the protection of unauthorised access to and damage to a computer and computer network, etc. It has a conditional safe harbour for an intermediary when it hosts or transmits third party content. Whether that safe harbor applies to entities that deploy autonomous agents that make independent decisions, rather than merely providing space on which to run user-supplied content, is open to interpretation. That's because the reason for the safe harbor is the intermediary's lack of knowledge and control over the particular content in issue, which conflicts with the idea that an agent should be free to act independently and with goal-directed behavior. Section 65B was added to the Indian Evidence Act, 1872 by the Information Technology Act, 2000. This section therefore regulates the use of electronic records in the Indian courts. In the case of Anvar P.V. the Supreme Court of India observed that when an electronic record is intended to be used as secondary evidence, it must be accompanied by a certificate under 65B(4) (identifying the electronic record, describing the manner of its production and confirmation that the device in relation to which the electronic record was produced was in proper working order at the relevant time). So, in this case the Supreme Court overturned a previous line of authority that was more liberal in its interpretation.

In Arjun Panditrao Khotkar vs. Kailash Kushanrao Gorantyal, the Court, later, reiterated and explained the issue to the effect that the certificate is necessary if the original device itself is not tendered as primary evidence. According to the proposed theory based on records produced by a computer that is being used regularly to store or process information for an activity regularly carried on, the ordinary functioning of the device and the record produced will have a stable, deterministic relationship. This is not a simple property to fulfill with autonomous agents, whose outputs are in fact probabilistic even when the inputs are materially identical.

Table 1: Comparative Overview of Legal and Regulatory Instruments Relevant to Forensic Readiness for Autonomous AI Agents

Instrument	Jurisdiction / Scope	Core Forensic-Relevant Obligation	Relevance and Limitation for Autonomous Agents
Budapest Convention on Cybercrime	Council of Europe (international, 2001/2004)	Criminal law harmonisation; procedural powers for preservation and production of stored data	Does not address AI agents specifically; provides the cross-border cooperation backbone

			applicable to agent-mediated evidence
Regulation (EU) 2024/1689 (Artificial Intelligence Act)	European Union (2024)	Article 12: mandatory automatic logging for high-risk AI systems; Article 26: deployer retention and cooperation duties	Closest existing model for AI-specific forensic readiness, but scope limited to systems classified as high-risk
General Data Protection Regulation (EU) 2016/679	European Union	Accountability principle; restrictions on solely automated decision-making with legal or similarly significant effects	Encourages record-keeping of human oversight but creates tension with data minimisation
Computer Fraud and Abuse Act, 18 U.S.C. 1030; Stored Communications Act, 18 U.S.C. 2701 et seq.	United States (federal)	Substantive cybercrime offences; framework for compelled disclosure of records held by service providers	Silent on autonomous agent conduct; interacts with Fourth Amendment limits recognised in Carpenter and Riley
Information Technology Act, 2000; Indian Evidence Act, 1872, s.65B	India	Substantive cybercrime offences; intermediary safe harbour; certification requirement for electronic evidence	s.65B certification model presumes deterministic record production, in tension with probabilistic agent outputs

Source: Compiled by the author from the primary instruments and judicial decisions discussed in Part V.

VI. EVIDENTIARY ADMISSIBILITY OF AI-MEDIATED AND AI-GENERATED RECORDS

As opposed to traditional computer-generated evidence, autonomous agents generate records which raise admissibility issues distinct from those raised by the record. Normally, the chain of custody is sorted by pinpointing few number of custodians who held the evidence at the time it was created until it was presented in court. If, however, a single agentic task creates records across a single model application programming interface (API) provider, cloud infrastructure host, one or more third-party tool provider and the deploying organisation, then the chain of custody question is significantly complicated. These agents have different retention periods on the documents and each can alter and remove evidence independently as per their policies. The law provided in the section 65B of Indian Evidence Act conditions the admissibility of a record as being produced by a computer in proper working order and used regularly in the activity. These frameworks were originally designed for systems whose output is reproducible if the same inputs were to be given to the computer and if it was operated properly. Generative models do not have this property in its strict sense, because there can be different reasoning processes and output from the same prompt. When courts or lawmakers face the scepticism of probabilistic variability when it comes to whether agents can generate admissible evidence, they will need to decide if this kind of variability will undermine the evidentiary reliability that these certification regimes are aiming to secure. Or perhaps the reliability is to be demonstrated by other evidence of the systems overall accuracy, its logging integrity and evidence of it not being tampered.

The notion recognised in Carpenter v. United States and Riley v. California that the Constitution is directly engaged. Requiring a full logging of agent behaviour as a forensic readiness regime suggests that an extensive record of the behaviour of all the people who interact with the agent will also be maintained. The reasoning of Carpenter – that a collection of digital records which itself is easily compiled is worthy of heightened privacy protection – would indicate that government access to agent logs, even when the private company is not a telecommunications provider, should be deemed constitutionally significant in doctrinally comparable jurisdictions. To be effective and resilient in law (e.g. against evidence admissibility issues), a forensic readiness framework must enable law enforcement access, using a warrant or other legally authorised means, not informally or administratively.

For another and primarily new evidentiary requirement there is the issue of log integrity. The evidence value of an agent logging its actions, either because it has been instructed to do so or under a compromise prompt, is dependent on the evidence that the logs have been produced and stored in a way that is not susceptible to retrospective modification.

Technically, this can be met by having cryptographically chained, append-only logging structures. These are the ones used in certificate transparency systems, and can be used to prove that the log entry was present in the log at that time, without any collusion or good faith on the part of the system creating the log.

International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 8
Aug 2026, PP 1-10

Table 2: Forensic Readiness Requirements Mapped to the Autonomous Agent Lifecycle

Lifecycle Stage	Forensic Readiness Requirement	Evidentiary Purpose	Primary Responsible Actor
Design and training	Baseline behavioural documentation; model and training data provenance records	Establishing an evidentiary baseline against which anomalous conduct can later be assessed	Model developer
Deployment and configuration	Records of tool access granted, permission scopes, and human oversight checkpoints configured	Enabling reconstruction of what the agent was authorised to do at the time of an incident	Deployer
Live operation	Cryptographically verifiable, append-only action and decision logs	Preserving a tamper-evident, chronological account of the agent's actual conduct	Deployer and tool/infrastructure providers
Incident detection and preservation	Expedited preservation requests; forensic imaging of relevant logs and system states	Preventing loss of volatile or short-retention data before formal legal process concludes	Deployer, investigators, and service providers
Post-incident investigation and disclosure	Certified extraction of records meeting applicable evidentiary certification requirements	Rendering the preserved records admissible before a court or tribunal	Investigating authority and deployer

Source: Compiled by the author, adapting the process stages recognised in ISO/IEC 27043:2015 and the ten-step process proposed by Rowlingson (2004) to the autonomous agent lifecycle.

VII. ACCOUNTABILITY MODELS FOR AUTONOMOUS AGENTS

When someone acts as an autonomous agent – an agent with no personal legal agency – then someone else must take responsibility for their actions. It can be accomplished by one or more of existing doctrinal paths. There are different implications for forensic readiness of each route. A product liability approach to an autonomous agent is to think of the agent as a product. The issue will be whether it harmed due to its defect in design, in its instructions or in its warnings. The onus will be on him to establish the actual operation of the system and its deviation from what it was reasonably expected to do. This will involve having access to design documentation, testing records and logs of incident. A vicarious liability similar to the employers liability rule would mean that the deploying organisation would be held liable for the agent's conduct where the conduct was engaged in for the purposes of the task the agent was deployed to perform. It would also remove the burden of the evidentiary process from the internal design of the authorisation to the scope of authorisation set up for the agent. A negligence approach would require one to consider whether the deployer has properly exercised reasonable care to supervise the agent. In particular: did the appropriate mechanisms of human oversight, which have already been foreshadowed in the EU AI Act (Article 14), exist and were operating at the time? Again, documentary evidence must be provided that oversight was actually provided and not just set up. There is no need for an inquiry into fault to be conducted under the strict liability approach, which aligns with the risk tiered logic that should be found in instruments like the Artificial Intelligence Act, in lower-risk systems. The same kind of inquiry would make the evidentiary burden easier for the petitioner, but the burden would be placed on the threshold classification objection to establish that the given agent is in scope of the regulation.

None of these models is exclusive of the others; and, like tort and criminal law, legal systems are likely to use some mix of these models, depending on the nature of the harm and the relationship between the parties. However, in each, the fact that the specific factual predicate which the particular model depends on exists is dependent on one or more other records that has to exist. It is therefore not surprising that forensic readiness is not a component of the accountability model, it is actually a pre-condition for any of them to do as intended. After all, a liability rule's doctrinal purity cannot be achieved without evidence, which means it's not a practical method of accountability.

Table 3: Comparative Accountability Models for Autonomous AI Agent Conduct

Model	Doctrinal Basis for Liability	Evidentiary Requirement	Likely Respondent
Product liability	Design, manufacturing, or instructional defect in the agent or its underlying model	Design documentation, pre-deployment testing records, and incident-specific logs	Model or agent developer
Vicarious liability	Conduct falling within the scope of tasks the agent was authorised to perform	Records of the permission scope and objectives configured for the agent	Deploying organisation

Negligence	Failure to exercise reasonable supervision or human oversight	Evidence of the oversight mechanisms actually implemented and exercised	Deployer or operator
Strict liability for high-risk systems	Membership of the agent within a regulator-defined high-risk category	Classification records and the automatic logs mandated for such systems	Provider or deployer, as allocated by the applicable regulation

Source: Compiled by the author, drawing on general principles of product liability, vicarious liability, negligence, and the risk-tiered structure of Regulation (EU) 2024/1689.

VIII. TOWARD A FORENSIC READINESS FRAMEWORK FOR AUTONOMOUS AGENTS

The paper proposes six elements, as critical to a forensic readiness framework adapted to autonomous agents, building on the process model of ISO/IEC 27043:2015 and the ten-step organisational programme suggested by Rowlingson. The minimum requirements for action-level logging are to capture the action target, the tools called, the input and output of each tool call, and any escalation to human review. Secondly, this is a record of action, not the model's reasoning trace. The disclosure of the internal reasoning trace raises new and legitimate trade secret and safety concerns that must be addressed by the forensic readiness regime, but cannot be ignored. Moreover, any logs of such actions would have to be cryptographically verifiable using append-only, hash-chained structures, so that the integrity of the logs could be shown without depending on continued cooperation of the system that created them. Independent agents to automate human interventions for consequence-generating actions help agencies to separate governance actions from choice. When it comes to tech companies using data subjects' personal data without permission, we know of bad experiences. Lastly, there should be harmonised retention periods, by contract or regulation, between different technical custodians who come into a particular deployment, whether it's the model provider, the cloud host, tool provider, and so on, ensuring that there are no gaps in the evidence simply because one technical custodian in the chain retains data for a shorter period than their counterpart. Fifth, mutual legal assistance and cross-border preservation mechanisms, as provided in the Budapest Convention, should be made more flexible to enable a request for a machine speed conduct. Even if they have a preservation request procedure with human speed conduct, it may be so voluminous and fast that it is impractical to respond in a timely fashion before the data becomes irrelevant or is gone. Sixth, human oversight should be planned to occur at specific, well-defined timepoints, both because the deployment of agents can only be made practical at specific timepoints, and because these are normative and legally defined moments in time where human decisions can be attributed.

This approach doesn't answer the question of who is responsible for this type of harm. This should be a matter of substantive law and the models of accountability that are referred to in Part VII. The function is limited but still very basic: to ensure that, regardless of which model of accountability that a legal system adopts, the record that is essential to apply that model is available at the time of the investigation or proceeding, and can be authenticated at that time.

Challenges and Limitations

There are several structural restrictions on the use of the above framework in practice. There is serious fragmentation of jurisdictions. An autonomous agent deployment can be made by a model provider in a jurisdiction, a cloud host in a second jurisdiction and end users in a third jurisdiction. The extension of cross-border cooperation is required in order for the investigating authority to have access to the data localisation, disclosure and privacy laws applicable to each of these relevant players, meaning that even a technically well-designed regime of reporting or logging may not be legally accessible to the investigating authority without requiring extended cross-border cooperation. Commercial confidentiality is yet another obstacle. This is a result of the interest of model developers and agent platform providers in safeguarding their model architecture and training processes. The requirement for logging internal reasoning and not external actions can lead to a misinterpretation of forensic readiness and the more general and more contentious transparency mandate. However, the vast number of logs and the rate at which they can be generated present a practical problem because a system that can take hundreds of discrete actions during one task will generate logs at a rate that may be too costly to store and too voluminous for an investigator to meaningfully examine. One danger here is that the logging becomes a mere formality, designed to comply with regulations—with no actual value for investigations.

Furthermore, limitations on the types of evidence that can be presented in court may also pose problems in obtaining data from private processes and communications. Forensic readiness regimes which require indefinite or long-term retention of fine-grained behavioural logs will be at odds with this principle. These will probably require specific retention periods, different access procedures and technical safeguards like pseudonymisation that allow for evidentiary usefulness but which restrict public access to the personal data. Lastly, the non-deterministic nature of the underlying models means that a comprehensive and unadulterated record of an agent's behaviour is not a complete explanation of why an agent acted as he did. This is because the specific factors by which a model had selected an action, over another, are not necessarily completely recoverable even from a complete record of its inputs and outputs. Therefore, the non-determinism characteristic of agentic forensic readiness is other than traditional forensic readiness which

is deterministic, a characteristic not thought of by current law.

Recommendations

The gap identified in this paper can be addressed through reforms at four different levels.

First, at the international level, countries that are parties to the Budapest Convention should assess whether the Convention's existing framework for expedited preservation and mutual legal assistance, together with its related protocols, is sufficient to deal with evidence generated by autonomous and semi-autonomous AI systems. These technologies operate at a speed and scale far beyond traditional human activity, which means that procedures originally designed for human-driven conduct may no longer be adequate. Additional guidance or even a dedicated protocol could help ensure that digital evidence from such systems can be effectively preserved and shared during cross-border investigations.

Second, regulatory authorities should consider expanding logging and record-keeping obligations beyond AI systems that are formally classified as high risk. Requirements similar to those contained in Article 12 of the European AI Act could also apply to autonomous agents that are capable of taking significant external actions without direct human oversight. From a forensic perspective, maintaining reliable records is important regardless of whether a system falls within a particular regulatory risk category. In addition, minimum retention periods for these records should be standardised across all technical custodians involved in deploying and operating an agent. Without consistent retention requirements, evidence may be lost simply because one participant in a multi-party technical ecosystem stores records for a shorter period than the others.

A third area for reform concerns technical standardisation. Organisations such as the International Organization for Standardization (ISO) could develop an extension or companion standard to ISO/IEC 27043:2015 that specifically addresses the forensic challenges posed by agentic AI systems. Such a standard could provide guidance on distinguishing action logs from internal reasoning traces, recommend cryptographic methods to preserve the integrity of logs across distributed environments, and establish interoperable logging formats that allow investigators to correlate evidence collected from multiple providers and platforms.

Finally, institutions responsible for investigation and adjudication must develop the expertise needed to deal with incidents involving autonomous agents. Law enforcement agencies should be trained to identify, obtain, and interpret agentic action logs, while courts need clearer guidance on the admissibility and evidentiary value of records produced by probabilistic AI systems. A useful starting point would be a certification framework comparable to that under Section 65B of the Indian Evidence Act, but adapted to account for the non-deterministic nature of AI-generated outputs. At the same time, specialised investigative units with technical expertise in agent orchestration platforms should be established so that complex AI-related cases can be handled more effectively.

IX. CONCLUSION

Autonomous AI agents introduce a fundamentally new element into the digital ecosystem. Although they are not legal actors, they operate between human intent and observable action, creating a layer of decision-making that existing cybercrime and digital evidence frameworks were never designed to address. As a result, traditional approaches to attribution, evidence collection, and accountability face challenges that did not exist in conventional computing environments.

The concept of forensic readiness remains valuable and continues to provide a strong foundation for digital investigations. However, it must be adapted to reflect the realities of agentic AI systems, which are characterised by probabilistic decision-making, shared technical custody across multiple entities, and the ability to perform large volumes of actions at remarkable speed. While the Budapest Convention has not yet been updated to specifically address these developments, it continues to offer an important framework for international cooperation in cyber investigations. Likewise, the EU AI Act provides the clearest current example of mandatory logging requirements, although those obligations are presently confined to a defined category of high-risk AI systems.

Comparative case law also offers useful guidance. Decisions such as Carpenter and Riley in the United States, together with Anvar P.V. v. P.K. Basheer and Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal in India, demonstrate the importance of protecting constitutional rights while maintaining rigorous standards for the authentication and admissibility of digital evidence. Any evidentiary framework for records generated by autonomous AI systems should be developed with these principles in mind.

Closing the accountability gap identified in this paper does not require an entirely new legal framework. Instead, it calls for the careful adaptation of established principles, including forensic readiness, contemporaneous logging, cryptographic integrity, clearly defined custodial responsibility, and effective cross-border cooperation. These concepts have already proven their value in traditional digital investigations and can be extended to address the unique challenges posed by agentic AI systems. Technical innovation should not become a barrier to legal accountability. Rather, legal and forensic frameworks must evolve alongside these technologies to ensure that responsibility can still be established when autonomous systems cause harm. The framework proposed in this paper is intended as a foundation for that process, one that can continue to develop as agentic AI technologies and their societal impacts become increasingly sophisticated.

Declarations

Ethics Approval and Consent to Participate

Not applicable. This literature does not involve any new studies with human or animal participants conducted by the authors.

Consent for Publication

Not applicable. The manuscript does not contain any person's data in any form.

Competing Interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

All authors contributed equally to the conceptualization, literature review, analysis, and writing of the manuscript. All authors read and approved the final version.

Acknowledgements

The authors thank the institutional support provided by SAM Global University, M.P., India, for facilitating this research.

REFERENCES

1. Statutes and Regulatory Instruments.

- Computer Fraud and Abuse Act, 18 U.S.C. 1030.
- Stored Communications Act, 18 U.S.C. 2701 et seq.
- Information Technology Act, 2000 (India).
- Indian Evidence Act, 1872, s. 65B (as inserted by the Information Technology Act, 2000).
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), OJ L 119.
- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act), OJ L.

2. International Instruments

- Council of Europe, Convention on Cybercrime (Budapest Convention), ETS No. 185, opened for signature 23 November 2001, entered into force 1 July 2004.

Case Law

India

- Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473.
- Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1.
- State (NCT of Delhi) v. Navjot Sandhu alias Afsan Guru, (2005) 11 SCC 600.
- United States
- Carpenter v. United States, 585 U.S. 296 (2018). .
- Riley v. California, 573 U.S. 373 (2014).
- United States v. Jones, 565 U.S. 400 (2012).

Books

- Casey, Eoghan, Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet (4th edn, Academic Press 2011).
- Carrier, Brian, File System Forensic Analysis (Addison-Wesley Professional 2005).
- Mason, Stephen and Daniel Seng (eds), Electronic Evidence and Electronic Signatures (6th edn, Institute of Advanced Legal Studies 2021).
- Nelson, Bill, Amelia Phillips and Christopher Steuart, Guide to Computer Forensics and Investigations (7th edn, Cengage Learning 2023).
- Rogers, Marcus and Matt Bishop (eds), The Science of Digital Forensics (Addison-Wesley Professional 2006).
- Bhardwaj, Lakshya, et al. Principles and Practices of Modern Forensic Science: A Comprehensive Textbook. 2026.

3. Journals

- Casey, Eoghan, 'Digital Evidence and the Crime Scene' (2004) 1(Suppl. 1) Digital Investigation S1–S6.
- Casey, Eoghan, 'What Does "Forensically Sound" Really Mean?' (2007) 4(2) Digital Investigation 49–50.
- Carrier, Brian and Eugene H. Spafford, 'Getting Physical with the Digital Investigation Process' (2003) 2(2) International Journal of Digital Evidence 1–20.
- Englbrecht, L., 'IT Forensic Readiness', in Sushil Jajodia, Pierangela Samarati and Moti Yung (eds), Encyclopedia of Cryptography, Security and Privacy (Springer 2025).
- Garfinkel, Simson L., 'Digital Forensics Research: The Next Ten Years' (2010) 7(Suppl. 1) Digital Investigation S64–S73.
- Horsman, Graeme, 'Formalising Investigative Processes within Digital Forensics' (2019) 30 Digital Investigation 10–16.
- Karie, Nickson M., Hossein S. Venter and Karim A. Elloumi, 'The State of the Art of Digital Forensic Investigation' (2019) 4(2) Forensic Sciences Research 63–79.
- Pollitt, Mark M., 'A History of Digital Forensics' (2010) 7(Suppl. 1) Digital Investigation S3–S10.
- Quick, Darren and Kim-Kwang Raymond Choo, 'Digital Forensic Intelligence: Data Subsets and Open Source Intelligence (DFINT+OSINT): A Timely and Cohesive Mix' (2015) 50 Future Generation Computer Systems 558–566.
- Rowlingson, Robert, 'A Ten Step Process for Forensic Readiness' (2004) 2(3) International Journal of Digital Evidence.

- Tan, John, 'Forensic Readiness' (2001) Information Security Technical Report 52–56.

.