STM JOURNALS

Journal of
Open Source Developments

ISSN: 2395-6704
Volume 13, Issue 1, 2026
January–April
DOI (Journal): 10.37591/JoOSD

https://journals.stmjournals.com/joosd

Review JoOSD

Open-Source Software Empowering Artificial Intelligence, Machine Learning, and Cyber Security: A Comprehensive Research Study

Sangeeta Singh*

Abstract

Open-source software (OSS) has become a foundational pillar for rapid innovation across artificial intelligence (AI), machine learning (ML), and cybersecurity. This paper delivers a comprehensive, journal-length analysis of OSS-driven ecosystems, emphasizing collaborative development, transparency, and accelerated deployment. By providing freely available libraries, tools, and frameworks, OSS makes it easier for developers and researchers to experiment, build models, and deploy solutions quickly. This study examines how OSS can be combined with AI and ML to improve cybersecurity, using a layered framework that covers data collection, processing, model training, threat detection, and automated response. Popular OSS tools are reviewed, showing how they improve detection accuracy, system efficiency, and scalability. While OSS offers many advantages, it also presents challenges such as exposed vulnerabilities, inconsistent updates, limited documentation, and integration difficulties. Strategies like secure coding, community oversight, and privacy-focused methods such as federated learning can help overcome these issues. The paper also discusses future directions, including explainable AI, autonomous security systems, and blockchain-based protection. Overall, OSS combined with AI and ML provides transparent, collaborative, and adaptable solutions that strengthen cybersecurity and support faster innovation across industries.

Keywords: Open-source software, artificial intelligence, machine learning, cybersecurity, deep learning, intrusion detection, open source intelligence (OSINT)

INTRODUCTION

In the modern digital era, technology is advancing rapidly, reshaping how organizations function, communicate, and secure their systems. One of the main forces behind this change is open-source software (OSS), which gives users the freedom to access, modify, and share software. This open approach has made OSS a key part of innovation, especially in fields such as artificial intelligence (AI), machine learning (ML), and cybersecurity [1].

AI and ML depend on data, algorithms, and computing tools, many of which are built and improved in open-source communities. These technologies allow systems to learn from data, recognize patterns, and make decisions with minimal human input. Simultaneously, the growing number and complexity of cyber threats have made cybersecurity more important than ever, pushing organizations to adopt smarter and faster security solutions [2–5].

Open-source technologies help to bring these areas together. Many modern cybersecurity systems use AI and ML models developed on OSS platforms

***Author For Correspondence**
Sangeeta Singh
E-mail: sangeeta25mu@gmail.com

Assistant Professor, Department of Computer Science Engineering, Madhav University, Pindwara, Rajasthan, India

Received Date: March 31, 2026
Accepted Date: April 17, 2026
Published Date: April 30, 2026

Citation: Sangeeta Singh. Open-Source Software Empowering Artificial Intelligence, Machine Learning, and Cyber Security: A Comprehensive Research Study. Journal of Open Source Developments. 2026; 13(1): 8–15p.

to process large amounts of data and detect unusual or harmful activities. However, the open nature of OSS can also create challenges, such as exposed vulnerabilities and the risk of misuse by attackers [6–9]. This study explores how OSS supports innovation in AI, ML, and cybersecurity, introduces a unified framework and system design, examines its advantages and limitations, and suggests future directions for safe and effective use (Table 1).

LITERATURE REVIEW

Prior research has consistently demonstrated a symbiotic relationship between OSS and AI-enabled cybersecurity. Surveys of AI for cybersecurity reports improved detection rates and reduced response times via supervised, unsupervised, and deep learning approaches [8, 4]. Foundational works on anomaly detection [5] and network intrusion detection [10] underscore the importance of data quality, feature engineering, and realistic evaluation.

Open-source AI introduces both transparency and risks. Studies have highlighted model inversion, data leakage, and adversarial manipulation as key concerns in open ecosystems [1]. Explainable AI (XAI) has emerged to address trust and interpretability challenges [11]. The expansion of OSINT has concurrently enhanced situational awareness in cyber defense [12, 13].

Recent preprints and applied studies have explored AI-driven forensic analysis, automated malware classification, and LLM-assisted threat intelligence. However, gaps remain in the standardized evaluation, secure supply chains, and governance of community-driven codebases (Table 2).

Table 1. Open-source software tools in AI/ML.

Tool name	Category	Use	Importance
TensorFlow	Deep learning framework	Used for building and training neural networks	Highly scalable, supports production-level AI models
PyTorch	Deep learning framework	Used for dynamic neural network development	Easy to learn, flexible, widely used in research
Scikit-learn	Machine learning library	Used for classification, regression, and clustering	Beginner-friendly and efficient for small datasets
Keras	High-level API	Simplifies deep learning model creation	Fast prototyping and user-friendly interface
Apache Spark MLlib	Big data ML	Used for distributed machine learning	Handles large-scale data efficiently
Hugging Face Transformers	NLP library	Used for text processing and language models	Pre-trained models reduce development time
OpenCV	Computer vision	Image and video processing	Real-time applications like face detection
Natural language toolkit (NLTK)	Natural language processing (NLP) toolkit	Text analysis and preprocessing	Good for educational and research purposes
Pandas	Data handling	Data manipulation and analysis	Essential for preprocessing structured data
NumPy	Numerical computing	Mathematical operations on arrays	Backbone of scientific computing in Python
Matplotlib	Data visualization	Plotting graphs and charts	Help with data understanding and insights
Seaborn	Data visualization	Advanced statistical visualizations	Improves data presentation quality
XGBoost	ML algorithm library	High-performance gradient boosting	Widely used in competitions and real-world tasks
LightGBM	ML algorithm library fast gradient boosting on large datasets	Efficient and faster than traditional methods	
FastAI	Deep learning library	Simplifies training in deep learning models	Speeds up development with less code

Table 2. Comparative analysis of literature review.

Author(s)	Year	Focus	Key insight
Chandola et al. [5]	2009	Anomaly detection	Taxonomy and benchmarks
Buczak and Guven [4]	2015	ML for IDS	Feature engineering critical
Sommer and Paxson [10]	2010	IDS realism	Closed-world limits
Kaur et al. [8]	2023	AI in cybersecurity	Automation benefits
Al-Kharusi et al. [1]	2024	OSS AI risk	Privacy and adversarial issues

AI, artificial intelligence; IDS, intrusion detection systems; ML, machine learning; OSS, open-source software

RESEARCH METHODOLOGY

This study follows a qualitative and analytical approach to understand the role of OSS in AI, ML, and cybersecurity. Instead of performing experiments, this study focuses on reviewing existing studies and comparing different methods and tools. A synthesis-based approach is used to combine insights from multiple sources, helping build a clear and balanced understanding of the topic [11, 13–15].

Data Sources

This study used reliable and well-recognized sources to maintain accuracy and credibility. Information was gathered from peer-reviewed journals available on platforms such as IEEE, ACM, Elsevier, and Springer. It also includes official guidelines, such as National Institute of Standards and Technology (NIST) standards and documentation from open-source projects. In addition, real-world case studies and industry reports were used to connect theory with practical applications [10, 16–18].

Analytical Approach

A structured method was used to examine the collected data. Important themes, such as system security, performance, and governance, were identified and studied. Different open-source tools and frameworks were compared to understand their advantages and limitations. Conceptual models are also created to explain system design and architecture simply and clearly [12, 19–21].

Evaluation Criteria

This study used key factors to evaluate the findings in a balanced manner. Accuracy was measured using metrics such as detection rate, precision, and recall. Scalability examines how systems manage increasing workloads. Transparency checks how easily results can be understood, while security focuses on potential risks, vulnerabilities, and update practices (Figure 1).

CONCEPTUAL FRAMEWORK (OSS–AI–CYBERSECURITY)

OSS Resource Layer

The OSS resource layer acts as the starting point of the framework by providing important resources, such as datasets, open libraries, and ready-to-use models. Because these resources are openly available, users can easily access and reuse them without building everything from the beginning. This encourages collaboration and saves both time and costs. This also makes advanced AI and cybersecurity tools more accessible to a wider group of users [22–24].

AI/ML Processing Layer

The AI/ML Processing Layer focuses on converting raw data into useful results. During training, the system learns patterns from the data, whereas validation checks its accuracy and helps to prevent errors. In the final stage, the trained model was used to make predictions on new data. This layer is essential for converting data into meaningful insights that support decision making [25].

Cybersecurity Analytics Layer

The Cybersecurity Analytics Layer is responsible for tracking and examining system activities to identify possible threats. It uses different tools and techniques to detect unusual behaviors and security risks. By analyzing patterns and spotting anomalies, this layer helps in the early detection and quick response to cyberattacks, thereby improving system safety [26].

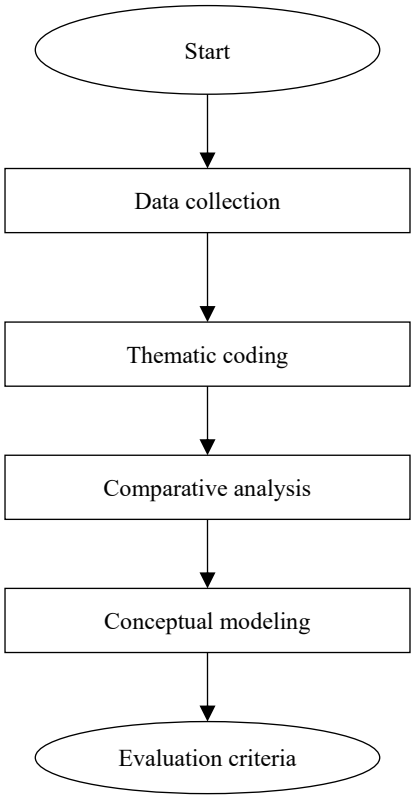

Figure 1. Research methodology flowchart.

Response and Governance Layer

The response and governance layer handles the system's reaction after a threat is detected. It supports automated actions to respond quickly and effectively. Security rules guide these actions to ensure proper control and compliance with the regulations. It also includes monitoring and review processes to maintain transparency and improve the system performance over time [27].

ROLE OF OPEN-SOURCE SOFTWARE IN AI AND ML

OSS plays an important role in the growth of AI and ML by offering shared tools and frameworks that make development faster and easier. It allows developers to experiment, test models, and deploy solutions without starting from scratch. Contributions from the global community help improve algorithms, test their performance, and keep systems updated. This open approach also supports transparency, making it easier to understand how models work and ensuring that results can be reproduced [28, 29].

1. Open-source platforms remove cost barriers, allowing anyone, from students to professionals, to use advanced AI and ML tools.
2. It promotes teamwork, where people collaborate, share knowledge, and improve technology.
3. Popular frameworks such as TensorFlow and PyTorch provide built-in functions that simplify complex model development tasks.
4. Developers can freely modify and customize the code to meet specific needs and create new solutions.
5. Transparency improves trust, as users can review the code to check for errors, bias, or ethical concerns.
6. It supports learning by giving beginners access to real tools and practical examples.
7. Open standards help different tools work together smoothly across platforms.
8. Development becomes quicker due to the availability of reusable components and libraries.
9. Active communities provide support, tutorials, and solutions to common problems.
10. Overall, OSS makes AI and ML more accessible, flexible, and continuously improving.

OPEN-SOURCE SOFTWARE IN CYBERSECURITY

OSS tools support continuous monitoring, in-depth analysis, and offensive and defensive security testing. They enable organizations to detect threats, assess vulnerabilities, and strengthen overall system protection (Table 3).

INTEGRATION OF AI, ML, AND OSS IN CYBERSECURITY

The combination of AI, ML, and OSS has made cybersecurity systems more advanced and effective. These technologies help identify threats quickly, automate responses, and improve overall system security.

Intrusion Detection Systems

Intrusion detection systems (IDS) monitor network activity to identify any unusual or suspicious behavior that could indicate an attack. They help organizations detect threats early and take action before damage occurs to the organization. By continuously studying traffic patterns, IDS improves system safety and reduces risks.

User and Entity Behavior Analytics

User and entity behavior analytics (UEBA) analyzes the behavior of users and devices within a system to detect unusual actions. It compares normal behavior with abnormal patterns to identify risks, such as insider threats or compromised accounts. This adds an extra layer of security by focusing on behavior rather than just system activity.

Malware and Ransomware Detection

Malware and ransomware detection systems use ML to identify harmful software. These systems recognize patterns linked to malicious programs and block them before they can spread. This helps protect the data and prevents system failure.

Threat Intelligence Systems

Threat intelligence systems gather and analyze information from different sources to understand possible security threats. These provide useful insights that help organizations prepare for and prevent cyberattacks. Real-time data helps make faster and better security decisions.

These applications use ML models to identify risks and predict threats before they occur, making cybersecurity systems more proactive and reliable.

CHALLENGES AND LIMITATIONS

Although OSS is widely used in cybersecurity, it has some drawbacks. Because the source code is openly available, it can sometimes expose weaknesses if not properly maintained. In addition, unclear documentation, uneven support from communities, and difficulties in combining different tools can reduce the efficiency and reliability of the system.

Table 3. Open-source software cybersecurity tools.

Tool	Type	Function and description
Snort	IDS	A signature-based intrusion detection system that monitors network traffic for known threats and generates alerts for suspicious activity.
Suricata	IDS/IPS	A high-performance network monitoring tool that can act as both an intrusion detection and prevention system, analyzing traffic in real time.
Wireshark	Analyzer	A powerful packet analyzer that captures and inspects network packets, helping diagnose network issues and detect anomalies.
Zeek	Network Analysis	A network security monitoring platform that focuses on behavioral analytics to identify unusual patterns and potential threats.
Metasploit	Pentest	A penetration testing framework that allows security professionals to simulate attacks, develop exploits, and evaluate system vulnerabilities.

IDS, Intrusion Detection System; IPS, Intrusion Prevention System.

Security Risks

Open-source systems may develop security issues if they are not updated or reviewed promptly. Hackers can exploit these weaknesses to breach systems or disrupt services. Regular updates and proper monitoring are necessary to avoid these risks.

Lack of Standardization

Open-source tools are often created by different developers using various methods. This can lead to compatibility problems when integrating multiple tools. Consequently, ensuring smooth system operation can be challenging.

Data Privacy Concerns

AI models often depend on large datasets that may contain private or sensitive information. If proper safeguards are not followed, these data can be exposed or misused. Protecting data through secure handling and anonymization is crucial.

Misuse of Technology

Powerful AI tools can also be used in harmful ways, such as creating advanced cyberattacks or malicious software. This makes it important to follow ethical practices and ensure the responsible use of technology.

Future Scope

The future of OSS in AI and cybersecurity offers many new possibilities for growth and improvement. There will be a stronger focus on building systems that are secure, reliable, and easy to understand and use. New approaches, such as privacy-focused learning, automated security solutions, and advanced data protection methods, will play an important role in keeping digital systems safe.

Secure and Trustworthy AI Frameworks

Future AI systems will be designed to be more secure and reliable. These frameworks will include strong protection measures to prevent misuse and to protect sensitive data. Ensuring accuracy, stability, and trust will be a major focus.

Explainable AI for Better Transparency

Explainable AI will help users understand how decisions are made by AI systems. Instead of unclear outputs, the systems provide simple explanations for their results. This will increase trust and help users make better decisions in the future.

Privacy-Preserving Techniques (Federated Learning)

Techniques such as federated learning allow systems to learn from data without directly sharing it. Data will remain on local devices, reducing the chances of exposure. This approach helps protect user privacy while improving model performance.

Autonomous Cybersecurity Systems

Future cybersecurity systems will become more automated, using AI to detect and respond to threats independently. These systems operate continuously and react quickly, reducing the need for manual intervention and improving security response times.

Blockchain-Based Data Protection

Blockchain technology will be used to store and manage data in a secure and decentralized manner. This helps prevent unauthorized changes and ensures data integrity. This makes it useful for protecting sensitive information in modern systems from adversarial attacks.

CONCLUSION

This study shows that OSS plays an important role in AI, ML, and cybersecurity. OSS allows

developers to collaborate, share resources, and build intelligent systems faster and more efficiently. By integrating OSS with AI and ML, organizations can improve threat detection, automate responses, and enhance overall system performance. It also helps reduce costs and makes systems more flexible and easier to understand. However, there are some challenges, such as open code risks, poor maintenance, and difficulty in using different tools together. These problems can be managed by following secure practices and improving community support. In the future, technologies like explainable AI, self-operating security systems, and blockchain will make these systems even stronger. In the end, OSS helps organizations build secure, scalable, and efficient systems. Its open and collaborative nature supports innovation and helps organizations stay prepared for new and growing cyber threats.

REFERENCES

1. Al-Kharusi Y, Khan A, Rizwan M, Bait-Suwailam MM. Open-source artificial intelligence privacy and security: a review. Computers. 2024;13(12):311. doi:10.3390/computers13120311.
2. Axelsson S. Research in intrusion-detection systems: a survey. Technical Report 98–17. Gothenburg: Department of Computer Engineering, Chalmers Tekniska Högskola; 1998.
3. Bishop CM, Nasrabadi NM. Pattern Recognition and Machine Learning. New York: Springer; 2006.
4. Buczak AL, Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun Surv Tutor. 2016;18(2):1153–1176. doi:10.1109/COMST.2015.2494502.
5. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv. 2009;41(3):1–58. doi:10.1145/1541880.1541882.
6. Jimmy F. Emerging threats: the latest cybersecurity risks and the role of artificial intelligence in enhancing cybersecurity defenses. Int J Sci Res Manag. 2021;9:564–574. doi:10.18535/ijsrm/v9i2.ec01.
7. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep Learning. Cambridge (MA): MIT Press; 2016.
8. Kaur R, Gabrijelčič D, Klobučar T. Artificial intelligence for cybersecurity: literature review and future research directions. Inf Fusion. 2023;97:101804. doi:10.1016/j.inffus.2023.101804.
9. Kim G, Lee S, Kim S. A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Syst Appl. 2014;41:1690–1700. doi:10.1016/j.eswa.2013.08.066.
10. Sommer R, Paxson V. Outside the closed world: on using machine learning for network intrusion detection. IEEE Symposium on Security and Privacy; 2010 May 16; Oakland, CA, USA. 2010. p. 305–316. doi:10.1109/SP.2010.25.
11. Mendes C, Rios TN. Explainable artificial intelligence and cybersecurity: a systematic literature review [preprint]. 2023. arXiv:2303.01259. doi:10.48550/arXiv.2303.01259.
12. Brilingaitė A, Bukauskas L, Juozapavičius A, Kutka E. Overcoming information-sharing challenges in cyber defence exercises. J Cybersecur. 2022;8(1):tyac001. doi:10.1093/cybsec/tyac001.
13. Szymoniak S, Foks K. Role of open source intelligence methods in cybersecurity. IEEE 17th International Scientific Conference on Informatics; 2024; Poprad, Slovakia. 2024. p. 390–395. doi:10.1109/Informatics62280.2024.10900930.
14. Sanz-Gómez M, Mayoral-Vilches V, Balassone F, Navarrete-Lozano LJ, Veas Chavez CRJ, del Mundo de Torres M. Cybersecurity AI benchmark (CAIBench): a meta-benchmark for evaluating cybersecurity AI agents [preprint]. 2025. arXiv:2510.24317. doi:10.48550/arXiv.2510.24317.
15. Mitchell TM. Does machine learning really work? AI Mag. 1997;18(3):11. doi:10.1609/aimag.v18i3.1303.
16. Ojha AK, Rao RA. The emergence of an organizational field: the case of open source software. Vikalpa. 2014;39(2):127–143. doi:10.1177/0256090920140212.
17. Sarker IH, Furhad MH, Nowrozy R. AI-driven cybersecurity: an overview, security intelligence modeling and research directions. SN Comput Sci. 2021;2(3):173. doi:10.1007/s42979-021-00557-0.
18. Scarfone K, Mell P. Guide to Intrusion Detection and Prevention Systems (IDPS). NIST Special Publication 800-94. Gaithersburg (MD): National Institute of Standards and Technology; 2007.

19. Stallman RM. Free Software, Free Society: Selected Essays of Richard M. Stallman. Morrisville (NC): Lulu Press; 2002.
20. Ma Y, Chen M. The human touch in AI: optimizing language learning through self-determination theory and teacher scaffolding. Front Psychol. 2025;16:1568239. doi:10.3389/fpsyg.2025.1568239. PubMed PMID: 40678438.
21. Bodenheim RC. Impact of the Shodan computer search engine on internet-facing industrial control system devices [master's thesis]. Wright-Patterson Air Force Base (OH): Air Force Institute of Technology, Air University; 2014.
22. Ishaq K, Javed HA. Implementing Snort intrusion prevention system (IPS) for network forensic analysis [preprint]. 2023. arXiv:2308.13589. doi:10.48550/arXiv.2308.13589.
23. Soepeno RAA. Wireshark: an effective tool for network analysis [course report]. Jakarta: Information Systems, Faculty of Engineering and Technology, Sampoerna University; Applied Computing, College of Applied Science and Technology, University of Arizona; 2023.
24. Ba Alawi Z. A comparative survey of PyTorch vs TensorFlow for deep learning: usability, performance, and deployment trade-offs [preprint]. 2025. arXiv:2508.04035. doi:10.48550/arXiv. 2508.04035.
25. Chaudhary A, Chouhan KS, Gajrani J, Sharma B. Deep learning with PyTorch. In: Mahrishi M, Hiran KK, Meena G, Sharma P, editors. Machine Learning and Deep learning in Real-Time Applications. Hershey (PA): IGI Global; 2020. p. 61–95. doi:10.4018/978-1-7998-3095-5.ch003.
26. Ajibode A, Bangash AA, Cogo FR, Adams B, Hassan AE. Towards semantic versioning of open pre-trained language model releases on Hugging Face. Empir Softw Eng. 2025;30(3):78. doi:10.1007/s10664-025-10631-3.
27. Verma A, Sankhyayan S, Jawanda K, Tandon S. Generative artificial intelligence and cybersecurity risks: issues and challenges. In: Fong S, Dey N, Joshi A, editors. ICT Analysis and Applications. ICT4SD 2024. Lecture Notes in Networks and Systems. Vol. 1161. Singapore: Springer; 2025. p. 363–375. doi:10.1007/978-981-97-8602-2_29.
28. Rahman MA, Francia G, Shahriar H, Cuzzocrea A, Mohamed A, Khan MU, et al. A survey of large language models (LLMs) for cybersecurity: opportunities and directions. 2025 IEEE International Conference on Big Data (BigData), Macau, China. 2025. p. 4333–4342. doi:10.1109/BigData66926.2025.11402639.
29. Naik B, Mehta A, Yagnik H, Shah M. The impacts of artificial intelligence techniques in augmentation of cybersecurity: a comprehensive review. Complex Intell Syst. 2022;8(2):1763–1780. doi:10.1007/s40747-021-00494-8.

