Annals of the „Ovidius" University of Constanța – Political Science Series
Analele Universității „Ovidius" din Constanța – Seria Științe Politice
Volume 14 (2025): 239-245
https://www.doi.org/10.61801/AUOC-SP.2025.11

TUOMO SIPOLA, JANNE ALATALO, MONIKA WOLFMAYR, AND TERO KOKKONEN, EDS., *ARTIFICIAL INTELLIGENCE FOR SECURITY: ENHANCING PROTECTION IN A CHANGING WORLD*, CHAM: SPRINGER NATURE, 2024, 307 P., ISBN 978-3031574511

Lokesh SWAMI*

Received: July 27th, 2025 Accepted for publication: October 1st, 2025

Abstract: This review critically examines *Artificial Intelligence for Security: Enhancing Protection in a Changing World* edited by Tuomo Sipola, Janne Alatalo, Monika Wolfmayr, and Tero Kokkonen, published in Springer Nature, in 2024. Structured in three parts: "Methodological Fundamentals", "Critical Infrastructure Protection", and "AI for Anomaly Detection", the volume brings together theoretical insights and applied studies to explore how AI enhances security systems. Part I engages with foundational concepts such as differential privacy, explainable AI, and adversarial robustness. Part II presents sector-specific applications ranging from logistics and smart grids to healthcare systems. Part III demonstrates AI's utility in real-time anomaly detection, providing empirical results on web-attack detection, log analysis, and Internet of Things (IoT) intrusion modeling. The book's strengths lie in its interdisciplinary approach, ethical framing, and strong emphasis on real-world applications. Case studies such as the use of fuzzy logic in smart grids and hybrid models for IoT defense underscore the book's practical relevance. However, limitations include a narrow domain scope (with less attention to areas like financial or defense security), minimal engagement with geopolitical dynamics, and overly technical vocabulary that may limit accessibility for non-specialist readers. Despite these constraints, the volume makes a substantial contribution to the field, integrating technical precision analysis with policy-aligned ethics, offering a valuable resource for cybersecurity researchers, practitioners, and policy architects concerned with building reliable AI-enabled systems.
Keywords: Artificial Intelligence, cybersecurity, critical infrastructure, anomaly detection, ethical AI

Rezumat: Această recenzie analizează critic volumul *Artificial Intelligence for Security: Enhancing Protection in a Changing World*, editat de Tuomo Sipola, Janne Alatalo, Monika Wolfmayr și Tero Kokkonen și publicat la Springer Nature, in 2024. Structurat în trei părți: „Fundamente Metodologice", „Protecția Infrastructurii Critice" și „Inteligența

* Lokesh Swami is an independent researcher based in Rajasthan, India, e-mail address: lokeshswami25@gmail.com.

Annals of the „Ovidius" University of Constanța – Political Science Series
Analele Universității „Ovidius" din Constanța – Seria Științe Politice
Volume 14 (2025): 239-245
https://www.doi.org/10.61801/AUOC-SP.2025.11

Artificială pentru Detectarea Anomaliilor", volumul combină perspective teoretice şi studii aplicate pentru a explora modul în care inteligenţa artificială (IA) contribuie la consolidarea securităţii. Prima parte abordează concepte fundamentale precum confidenţialitatea diferenţială, inteligenţa artificială explicabilă şi robusteţea împotriva atacurilor adversariale. Partea a doua prezintă aplicaţii sectoriale din domenii precum logistică, reţele inteligente şi sistemele de sănătate. Partea a treia evidenţiază utilitatea IA în detectarea anomaliilor în timp real, prezentând rezultate empirice cu privire la detectarea atacurilor web, analiza jurnalelor de sistem şi modelarea intruziunilor în în Internetul Lucrurilor (IoT). Punctele forte ale cărţii includ abordarea interdisciplinară, accentul pus pe etica IA-ului şi accentul pus pe dimesiunea aplicativă. Studiile de caz, precum utilizarea logicii fuzzy în reţele inteligente şi a modelelor hibride pentru apărarea IoT, evidenţiază relevanţa practică a volumului. Cu toate acestea, limitările constau în sfera de aplicare restrânsă a domeniului (acordând mai puţină atenţie asupra securităţii financiare sau militare), observaţii minime cu privier la dinamici geopolitice şi vocabularul tehnic predominant, care poate îngreuna accesibilitatea pentru cititorii nespecializaţi. În ciuda acestor constrângeri, volumul oferă o contribuţie semnificativă domeniului, integrand analiza tehnică de precizie cu etica aliniată la politici, constituind o resursă valoroasă pentru cercetători, practicieni şi factori de decizie preocupaţi de dezvoltarea unor sisteme IA fiabile în domeniul securităţii cibernetice.
Cuvinte cheie: inteligență artificială, securitate cibernetică, infrastructură critică, detectarea anomaliilor, etică IA

The book is structured to guide the reader from foundational AI principles to concrete applications in security-critical contexts. According to the editors' Preface, Part I "Methodological Fundamentals of Artificial Intelligence" introduces the general frameworks and main challenges in AI security. It opens with broad discussions (chapters by Adrowitzer et al. provide "a blueprint towards a safe world with AI and its development and application" (p. v), while Holmström et al. analyze AI in organizational/managerial cybersecurity, questioning its "silver bullet" hype). The middle chapters of Part I delve into core technical topics: Minna Kilpala and Tommi Kärkkäinen survey differential privacy models, while Sarah van Gerwen and colleagues address explainable AI in threat intelligence. Janševskis and Ošis examine secure knowledge discovery frameworks, emphasizing that data and knowledge extraction must include security considerations. Finally, Glazunov and Zarras's chapter concludes Part I with a detailed analysis of deep learning robustness: they review various attack strategies on neural networks and assess available defenses, grounding the volume in current adversarial Machine Learning (ML) research.

Part II, "AI for Critical Infrastructure Protection", focuses on sector-specific applications. The chapter by Nweke and Yayilgan frames AI's role in securing cyber-physical systems. The volume then includes domain-specific studies: Kai Rasmus discusses AI in small and medium-sized enterprises; Mikko Kiviharju's chapter addresses cybersecurity in logistics; Martinelli, Mercaldo, and

Volume 14 (2025): 239-245
https://www.doi.org/10.61801/AUOC-SP.2025.11

Santone develop fuzzy-logic-based machine learning for smart grid stability; Zolotukhin, Zhang, and Hämäläinen examine adversarial examples in mobile communication networks; while Jonske et al. explore teaching machine learning operating with medical / healthcare data. These contributions collectively demonstrate how AI can reinforce resilience in vital sectors (e.g. transportation, energy, communications, healthcare) while also introducing new challenges for each of the domains analyzed.

Part III, "AI for Anomaly Detection", presents real-world detection systems. Simone Falzone, Gabriele Gühring, and Benjamin Jung's chapter examines automated monitoring of log data, showing how machine learning can identify abnormal events in system logs in real time. Pojan Shahrivar and Stuart Millar focus on web applications, noting the surge in web attacks. They design an ML-based detector to flag malicious scanning. Finally, Mansour Alqarni and Akramul Azim address the issues surrounding IoT security by building a hybrid model that captures both spatial and temporal patterns. They explain that individually Convolutional Neural Networks (CNNs) (for spatial features) and Recurrent Neural Networks (RNNs) (for temporal sequences) each miss aspects of IoT traffic, so their hybrid model integrates both modalities. As the authors observe, this integration is "a crucial step toward more effective and adaptable IoT intrusion detection" (p. 350).

Overall, the book's three parts, from fundamentals to application and case studies, provides the reader with a well-thought and compelling narrative. It mirrors the editors' intent to balance conceptual discussions with applied examples. Each part includes both chapters reviewing literature or analytical frameworks and empirical studies, presenting experiments and results. For instance, in the anomaly-detection section the three chapters include experimental results on real log or network data, illustrating practical AI deployment. In this way, the volume interweaves theory and practice. Its interdisciplinary outlook is evident in Part I (which bridges computer science, data science, and management theory) as well as in the diversity of Part II (engineering, industry, and even educational perspectives on AI). The editors have curated contributions from academia and industry alike that enrich the book's outlook. The breadth of topics (from ethical AI design to smart grid monitoring) demonstrates the book's intention to cover the multifaceted role of AI in the modern security landscape.

In terms of the strengths of the book, I would note several aspects. First, the interdisciplinary and holistic approach ensures that the book's integrates technical depth with broader security concerns. The editors explicitly combine computer science, engineering, and social-science perspectives, for example, Part I pairs algorithmic chapters (differential privacy, deep learning robustness) with managerial or organizational analyses (Holmström et al. on the "silver bullet" myth in cybersecurity). Similarly, chapters on explainability and trust make connections to ethics. This interdisciplinary methodology provides a

Volume 14 (2025): 239-245
https://www.doi.org/10.61801/AUOC-SP.2025.11

more comprehensive understanding. As one contribution notes, addressing Explainable Artificial Intelligence (XAI) can "allow stakeholders to trust AI systems, detect biases, and identify potential vulnerabilities" (p. 18), directly linking technical explainability to social trust. The volume thus offers a comprehensive perspective: it does not treat AI security purely as an engineering problem, but also as a socio-technical issue.

A second strength derives from the book's emphasis on policy ethics as ethical AI concepts are woven throughout the text. Differential Privacy (DP), for example, is given a thorough treatment by Kilpala and Kärkkäinen: they note that DP's rigorous definition "ensures that personalized information remains non-disclosed", highlighting AI's privacy safeguards. The Explainable AI chapter explicitly frames XAI as a means to generate human-understandable accounts of AI decisions, meanwhile, in the broader organizational context, Holmström et al. discuss information classification and accountability. By foregrounding these topics, the editors align the book with current global ethics guidelines. Transparency, accountability, fairness and trust – key issues in AI governance – are also given attention. For example, Adrowitzer et al.'s chapter emphasizes that lifting "the veil on the black box" helps build "more ethical, trustworthy, and responsible AI systems" (p. 19) that benefit society. In this regard, the book can also serve to inform policymakers: it highlights how DP and XAI as practical mechanisms support legal and normative frameworks (e.g. data protection laws, standards for trustworthy AI, etc.).

Thirdly, the volume provides several case studies that have practical relevance with many chapters presenting detailed examples or experiments. For instance, Shahrivar and Millar's project on web-application scanning is grounded in real production data. The authors point out that in a "production environment with multiple apps and millions of events, it is not feasible to check each [alert] by hand" (p. 326). This real-world constraint motivates their ML solution while their results – an ML classifier detecting Dynamic Application Security Testing (DAST) attacks – demonstrate practical success. Likewise, Falzone et al. empirically show how preprocessing steps influence the performance of Principal Component Analysis (PCA), clustering, and deep-learning algorithms in log anomaly detection. Other studies (e.g. Kiviharju on logistics cybersecurity) include surveys of industry threats and standards. These applications underscore that the book is not merely theoretical: it provides concrete, actionable insights for practitioners. For example, Martinelli, Mercaldo, and Santone's approach explicitly seeks interpretability. As the authors note, their model incorporates "explainability, aimed to understand how the model is working from a global point of view" (p. 222), illustrating how practitioners can combine AI with domain expertise.

Fourthly, another strength is found in the book's clear and coherent structure. The chapters frequently review related work and define terms, for example, Glazunov and Zarras's chapter on deep nets begins by summarizing

Volume 14 (2025): 239-245
https://www.doi.org/10.61801/AUOC-SP.2025.11

known Deep Neural Network (DNN) attacks and defenses, setting the stage for new results. Similarly, the introductory sections of many chapters review background (the Jonske et al. chapter starts by tracing AI's history since the 1950s, for instance). This explanative approach aids comprehension since even highly technical chapters frame their contributions in understandable ways, while sections often end with conclusions or future directions.

Lastly, on the topic of strengths, I would also note the volume's relevance to diverse audiences. By explicitly addressing ethics, technology, and case studies, the book appeals to scholars, industry experts, and policy stakeholders. It positions itself as a resource for anyone concerned with modern cybersecurity and AI. As the editors state, "Understanding latest advancements in this field should be useful to ... experts ... who want to follow research and the latest trends." (p. vi). For example, academic researchers can appreciate the thorough literature reviews, while security professionals can learn from the anomaly-detection techniques. Policymakers and regulators can find value in the chapters on privacy, trust, and critical infrastructure resilience. The volume thus succeeds in bridging multiple communities: it translates cutting-edge AI research into insights on "build(ing) more secure systems" (p. vi) in practice.

On the issue of limitations, I would argue that a major blind spot is the limited insights from a geopolitical and global context. The book largely treats AI security in a technical and organizational vacuum, without properly engaging with geopolitical or cross-cultural dimensions. Except for passing mentions (e.g. data sovereignty), the chapters do not analyze how national policies, international rivalries, or cultural differences shape AI security. In an era of AI arms races and global digital governance (e.g. AI regulations by the EU, US, China), a discussion of how these dynamics influence AI adoption in security would have strengthened the volume. For example, critical infrastructure protection often depends on government strategy and international standards, but such perspectives are not fully developed. Expanding this dimension could have provided a more holistic view of AI's global impact.

On a different level, another shortcoming focuses on the limited accessibility for non-specialists. While the technical depth is a strength, it also poses a challenge given that some chapters assume familiarity with advanced AI concepts and use dense jargon. Readers without a strong machine learning background might struggle with sections on, for example, model extraction attacks, or the mathematical formalism of differential privacy. The volume lacks auxiliary tools (like a glossary of terms or appendices) that could help lay readers. Although this level of detail is expected in a research-oriented book, adding more explanatory context or simplifying some discussions could broaden the audience considering how AI has become increasingly integrated in day-to-day activities.

At the interdisciplinary level though the book is interdisciplinary in nature, the connections between technical and social aspects are sometimes

Volume 14 (2025): 239-245
https://www.doi.org/10.61801/AUOC-SP.2025.11

implicit rather than explicit. Each chapter addresses its niche well, but overarching synthesis is not explicitly drawn. A concluding chapter synthesizing insights across chapters (for instance, comparing how DP and XAI together contribute to "trustworthy AI") might have better unified the book's contributions. Moreover, while certain topics are comprehensively covered (logistics, smart grids, mobile networks, IoT), other important security contexts (e.g. cloud/edge computing, defense applications, financial system security) receive little or no attention. Similarly, while ethical issues are discussed, topics like algorithmic bias or social impact (beyond privacy and transparency) could have been emphasized more and broadened the ethical discussion.

In terms of broader implications, *Artificial Intelligence for Security*'s scope extends beyond its immediate case studies. Foremost, it underscores that AI is not just a set of algorithms, but a transformative factor in societal resilience. By framing AI security as critical to infrastructure protection, the volume acts as a call to use AI responsibly in an interconnected world. For example, emphasizing the safeguarding of energy grids and transportation nodes highlight that AI failure or misuse could have cascading effects on societies. As Falzone et al. note, modern systems generate "an ever-increasing number of log files" (p. 295) making human monitoring nearly impossible, which implies that AI-based solutions are not optional but necessary for early threat detection. Thus, a point could be made that investing in AI security research is essential for public welfare.

Furthermore, the dual focus on security and ethics helps shape the discourse on "trustworthy AI". Through the chapters on differential privacy and explainability, the book contributes to the policy discussion about how to make AI systems transparent and accountable. The volume's coverage of these topics aligns with ongoing global efforts (e.g. EU's AI Act, Institute of Electrical and Electronics Engineers' (IEEE) AI ethics standards) to regulate AI. For instance, referencing DP and XAI in a security context shows that privacy and transparency can be built into defensive technologies, not just compliance checkboxes. This integrated approach may influence both the IT community (to prioritize interpretable models) and regulators (to recognize technical solutions).

The book also bridges the gap between research and practice. The detailed case studies and results can inform practitioners on best practices to gain insight into current real-world problems – they can see what data and metrics matter in industrial settings (as shown in the log-monitoring chapter).

Finally, by highlighting remaining challenges, the book sets an agenda for future directions of research. The concluding sections of several chapters point to open problems: for example, Alqarni and Azim stress the difficulty of obtaining balanced IoT datasets and detecting rare attacks. These threads suggest directions for researchers (e.g. better data collection). Policymakers may also glean from the case studies that cybersecurity budgets and standards must adapt to AI-enabled systems. In all, the volume positions AI as a tool that can

Volume 14 (2025): 239-245
https://www.doi.org/10.61801/AUOC-SP.2025.11

both strengthen security and create new security needs. The volume's balanced treatment implies that stakeholders must advance AI innovations hand-in-hand with governance measures to maximize benefits and minimize harm.

In conclusion, *Artificial Intelligence for Security* is a timely and substantive contribution to the emerging literature on AI's role in cybersecurity and infrastructure protection. Its greatest achievement is the balance it strikes between advanced technical content and the depth of reflection. The editors and contributors provide a comprehensive resource that spans privacy-preserving analytics, system robustness, and practical detection tools. The volume's multidisciplinary outlook and emphasis on ethical guardrails make it all the more valuable as it demonstrates how AI research can be aligned with societal and governance concerns.

Nevertheless, the review's critique stands, future work should incorporate a wider geopolitical perspective. AI security does not exist in a vacuum, and addressing global policy, cultural, and legal dimensions would enhance the policymakers' ability to regulate the industry.

As a final observation, I would posit that the book will serve as a useful guide for scholars charting new research, for practitioners seeking AI solutions to real problems, and for policymakers looking for technically grounded insights. Ultimately, the editors succeed in delivering "unique perspectives to enhancing protection" (p. 6) in an AI-driven era, with the book representing as a significant step toward understanding how intelligent systems can help safeguard the future – a future in which security and innovation must advance together.