Politécnica

Revista Científica

Volume 32, dezembro de 2025

Cybersecurity & Auditing

Orgs.

José Carlos Morais
Mário Dias Lousã

Ficha Técnica

Diretor	Jorge Simões	
Subdiretor	Fernando Almeida	
Editor	ISPGAYA – Instituto Superior Politécnico Gaya	
Corpo Editorial	Ana Paula Cabral José Carlos Morais José Duarte Santos Mário Dias Lousã	
Comissão Científica	Ana Húngara (ISPGAYA, Portugal) António Gregório (ISPGAYA, Portugal) Arianne Pereira (ISPGAYA, Portugal) Carlos Bessa (ISPGAYA, Portugal) Carlos Costa (U. de Aveiro, Portugal) Catarina Lucas (ISPGAYA, Portugal) Dorothy Bedford (U. Roehampton, R. U.) Eric Zanghi (ISPGAYA, Portugal) Fernando Almeida (ISPGAYA, Portugal) Filipa Silva (ISPGAYA, Portugal) Filipe Silva (ISPGAYA, Portugal) Joana Becker Paulo (ISPGAYA, Portugal) João Álvaro Carvalho (U. Minho, Portugal) João Rocha Monteiro (ISPGAYA, Portugal)	Tiago Costa (ISPGAYA, Portugal) Jorge Simões (ISPGAYA, Portugal) José Candeias Filipe (ISCTE - IUL, Portugal) José Carlos Morais (ISPGAYA, Portugal) José Castro Oliveira (ISPGAYA, Portugal) José Noronha (ISPGAYA, Portugal) José Tavares (U. Aveiro, Portugal) Lídia Carvalho (ISPGAYA, Portugal) Maciel Barbosa (U. Porto, Portugal) Manuel Nogueira (ISPGAYA, Portugal) Maria Manuela Maia (ISPGAYA, Portugal) Óscar Lima da Silva (ISPGAYA, Portugal) Sérgio Sargo Lopes (ISPGAYA, Portugal) Vasconcelos Raposo, (UTAD, Portugal)
Coordenação e Revisão Editorial	José Carlos Morais	
Administração Redação Sede do Editor	Instituto Superior Politécnico Gaya Avenida dos Descobrimentos, 333 4400-103 Vila Nova de Gaia www.ispgaya.pt	
Propriedade	C.E.P. - Cooperativa de Ensino Politécnico CRL	
NIPC	NIPC 501962433	
Contacto	Tel. 22 374 57 30/3 Fax 22 374 57 39	
ISSN	0874-8799	
Nº ERC	123623	
Depósito Legal	153750/00	
ISBN – nº especial	978-972-8182-26-7	
DOI	https://doi.org/10.58086/pd1z-7855	
Frequência	Anual	
n.º	Vol. XXXII - Dezembro de 2025	

Estatuto Editorial: https://comum.rcaap.pt/retrieve/95533/Estatuto%20editorial%202020%20Timbrado.pdf

Índice

EDITORIAL

Os desenvolvimentos tecnológicos tiveram, desde sempre, maior ou menor impacto nas sociedades humanas que lhes deram origem. Este facto não constitui nenhuma novidade. A humanidade já lidou com esses impactos de forma mais evidente, pelo menos desde a Revolução Industrial.

Mais recentemente, na viragem para o século XXI, o uso generalizado de computadores e o fácil acesso a grandes volumes de informação, proporcionado pelas tecnologias associadas à Internet, tiveram implicações globais, nomeadamente na vida social, na forma como se trabalha e nas formas como se ensina e se aprende, para referir apenas algumas áreas.

Aquilo que atualmente faz a diferença é a velocidade com que os novos desenvolvimentos tecnológicos atingem as sociedades humanas e o potencial que têm para alterar profundamente o seu funcionamento e as suas regras. Também, tal como sempre aconteceu, a tecnologia pode, por um lado, ser usada para beneficiar a humanidade como um todo e, por outro, pode ser utilizada para favorecer apenas alguns indivíduos em detrimento de outros, servindo interesses políticos ou económicos ou facilitando ações criminosas. Quanto maior o poder das tecnologias, maiores serão os seus efeitos benéficos, mas também maiores serão os seus perigos potenciais.

No mundo atual, já fortemente dependente das tecnologias de informação e de quem as controla, os desenvolvimentos registados na última década na área da Inteligência Artificial — que permitiram o fácil acesso a uma panóplia de ferramentas e tecnologias associadas — são um exemplo evidente. A Inteligência Artificial integra um conjunto de tecnologias e ferramentas, com destaque para os grandes modelos de linguagem (LLM), que causaram forte impacto num curto espaço de tempo, com implicações em todas as áreas da sociedade e com um potencial de evolução difícil de antecipar.

Neste momento, em que grande parte das atividades humanas depende, de alguma forma, de ferramentas poderosas cujo funcionamento ultrapassa a compreensão da maioria dos seus utilizadores, a necessidade de um contexto que promova o uso ético, responsável e seguro dessas ferramentas é fundamental para a manutenção da coesão social. Neste contexto, a cibersegurança assume um papel essencial. A necessidade de garantir a autenticidade, a integridade, a disponibilidade e a confidencialidade da informação, a par de mecanismos de auditoria, certificação e atribuição de responsabilidades relativas a essa informação, é crucial

em sociedades cada vez mais digitalizadas. Mais uma vez, a Inteligência Artificial tem o potencial de aumentar as ameaças, mas também de contribuir para mitigar riscos e reduzir vulnerabilidades.

Neste volume, é possível encontrar diversas vertentes da cibersegurança em diferentes contextos. São apresentados estudos bibliométricos que incidem sobre vulnerabilidades no desenvolvimento e uso de aplicações informáticas, acesso a dados sensíveis e segurança de redes informáticas associada ao trabalho remoto. Outros estudos da mesma natureza, de forma transversal, abordam o uso ético e responsável da Inteligência Artificial e em apoio à cibersegurança. Há ainda referências a tecnologias emergentes, como blockchain e a Internet das Coisas (IoT), sendo também abordado o tema do cibercrime.

Os artigos aqui apresentados, cobrindo uma variedade de temas e tecnologias, tendo a cibersegurança como elemento comum, constituem um contributo relevante para esta área. Ao revelarem e analisarem produção científica atual e pertinente, permitem conhecer melhor o estado da arte da investigação em cibersegurança.

Jorge Simões

Janeiro 2026

https://doi.org/10.58086/2q80-9428

SECURITY AUTOMATION AND VULNERABILITY ANALYSIS IN .NET APPLICATIONS: A BIBLIOMETRIC REVIEW

Ruben Pinto[1]✉, **Mário Dias Lousã**[1,2], **José Carlos Morais**[1,3]

[1] Instituto Superior Politécnico Gaya (ISPGAYA), Portugal.
[2] Insight - Piaget Research Center for Ecological Human Developement, Portugal.
[3] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal.
✉ Corresponding authors**:** ispg2021103030@ispgaya.pt

Abstract

Application security is a crucial requirement in modern software, but manually detecting vulnerabilities is time-consuming and prone to failure. In the context of .NET platforms, widely used in corporate development, automation of security checks and vulnerability analysis emerges as a promising approach to efficiently mitigate risks. This article presents a bibliometric review on "Security Automation and Vulnerability Analysis in .NET Applications", identifying research trends and existing gaps. Publications indexed in The Lens database (2004–2025) were analyzed following PRISMA criteria, complemented by bibliometric techniques (co-authorship, co-citation, and co-occurrence of keywords) using tools such as VOSviewer. The results reveal a significant growth in work in the last decade, mainly addressing web vulnerabilities (e.g., SQL injection and XSS), as well as recent approaches to machine learning. However, there are important gaps, including the scarcity of studies specifically focused on the .NET ecosystem and low levels of collaboration among researchers. In short, although security automation has advanced, there are still research opportunities to fill the identified gaps, namely by adapting and expanding techniques for the .NET context.

Keywords: DevSecOps, Static Analysis (SAST); ASP:NET; Security Vulnerability Detection; Vulnerability analysis; Bibliometric review.

1. Introduction

The increasing digitization of services and the growing complexity of web architecture has intensified the need for effective application protection mechanisms, especially in corporate environments based on frameworks such as .NET. Vulnerabilities such as SQL injection,

XSS, and authentication flaws continue to be among the most exploited by malicious actors (OWASP, 2023), requiring systematized prevention approaches. In this context, security automation, notably through techniques such as Static Application Security Testing (SAST) and integrations into DevSecOps pipelines, has gained prominence (Williams & Wichers, 2021).

However, a fragmentation is observed in the scientific body on vulnerability analysis in .NET environments. Several studies address isolated technical aspects, but few provide a panoramic view of the state of the art and emerging trends in this intersection. Thus, the following research question arises: What are the gaps, trends, and collaborative patterns in the scientific production on security automation and vulnerability analysis in applications developed with the .NET framework?

To answer this question, a bibliometric review was used, since this method allows not only the synthesis of the knowledge produced, but also the objective and quantitative identification of co-authorship patterns, recurrence of themes, and temporal evolution of publications (Donthu et al., 2021). Unlike the traditional systematic review, which focuses on the qualitative and in-depth analysis of a limited subset of studies, bibliometrics enables a structural mapping of all scientific production in each domain, based on rigorous metrics and network visualizations (Zupic & Čater, 2015).

This article aims to present a comprehensive overview of scientific production related to automated security in .NET applications, identifying the main topics, collaborations between authors, existing gaps, and opportunities for future research. The document is organized as follows: section two presents the theoretical framework; section three describes the bibliometric methodology adopted; section four presents the results; section five critically discusses the findings; and section six concludes with conclusions and suggestions for future studies.

2. Literature Review and Conceptualization

The increasing sophistication of cyberattacks, coupled with the rapid evolution of corporate web architecture, has led to the recognition of security as a structural requirement in software development. In this scenario, security automation has gained prominence as an essential strategy for mitigating risks continuously and integrated into the development process

(Williams & Wichers, 2021). Concepts such as DevSecOps promote the inclusion of security practices from the early stages of the software lifecycle, focusing on automated testing and early vulnerability detection.

In the technical context, static analysis (Static Application Security Testing – SAST) and dynamic analysis (DAST) tools play central roles in this effort. Static analysis allows the identification of vulnerabilities in the source code without the need to execute the application, which facilitates integration into CI/CD pipelines (Sharma & Sheth, 2017). The literature demonstrates that such practices significantly reduce the cost of remediation and increase preventive security coverage (Kausar et al., 2019). Despite these advances, a notable gap is observed in scientific production focused on applications developed with the .NET framework, one of the most widely used in enterprise software development. Much of the literature addresses generic concepts of web application security, but there is a scarcity of studies that explore in depth the particularities of the .NET ecosystem, such as the use of ASP.NET, C#, or the specifics of the CLR (Common Language Runtime).

In this context, bibliometric analysis emerges as an appropriate methodology for mapping and understanding the state of the art. It is a quantitative approach that allows the identification of patterns in scientific production, co-authorship networks, recurring themes, and emerging trends, based on large volumes of publications (Donthu et al., 2021). Unlike traditional systematic reviews, which aim to synthesize empirical results based on a restricted sample, bibliometrics provides a broad structural view, allowing the visual and measurable identification of areas of concentration and thematic gaps (Zupic & Čater, 2015). By applying this type of analysis to the area of automated security in .NET, it becomes possible to identify the most active authors, the most frequently addressed topics, the preferred publication channels, as well as the gaps in scientific production. This systematic understanding constitutes a solid basis for guiding future research and supporting strategic decisions in the secure software industry.

3. Methodology

This study followed a bibliometric approach, according to the methodological guidelines proposed by Donthu et al. (2021), with visual and analytical support from tools such as VOSviewer for network analysis (Van Eck & Waltman, 2010). To ensure transparency and

reproducibility, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol, frequently applied in systematic reviews and adapted here to the bibliometric context, was also adopted (Page et al., 2021).

3.1. Data source and search strategy

The selected database was The Lens (www.lens.org) It offers broad coverage of scientific and technical literature, as well as robust export resources in structured bibliographic format. The search was conducted in December 2025, with the following keywords applied to the "title, abstract and keywords" field:

("security automation" OR "automated security" OR "vulnerability detection" OR "static analysis" OR "SAST") AND (".NET" OR "ASP.NET")

Documents published between 2012 and 2025 were considered, with languages limited to English and Portuguese, and publication types restricted to peer-reviewed journal articles and conference proceedings.

3.2 Inclusion and Exclusion Criteria

The inclusion criteria applied were:

- Posts related to automated security, vulnerability analysis, or DevSecOps practices in .NET or ASP.NET applications;
- Studies using techniques such as SAST, DAST, or Machine Learning in vulnerability detection.
- Works with full-text access.

The exclusion criteria were:

- Studies focused on other platforms (Java, PHP, etc.) unrelated to .NET;
- Duplicate documents;
- Technical reports, books, dissertations, or non-peer-reviewed communications;
- Works without relevant data on security automation.

3.3. Selection Process (PRISMA)

The selection process followed four stages, according to the PRISMA model (Page et al., 2021):

- Identification: 768 documents were initially retrieved from The Lens database.
- Screening: After removing duplicates and reading titles and abstracts, 412 articles were retained.
- Eligibility: After full reading, 167 articles were excluded for not meeting the criteria.
- Inclusion: The final corpus consisted of 245 scientific articles, which were used for bibliometric analyses.

Figure 1 presents the PRISMA diagram with the document inclusion and exclusion flow.

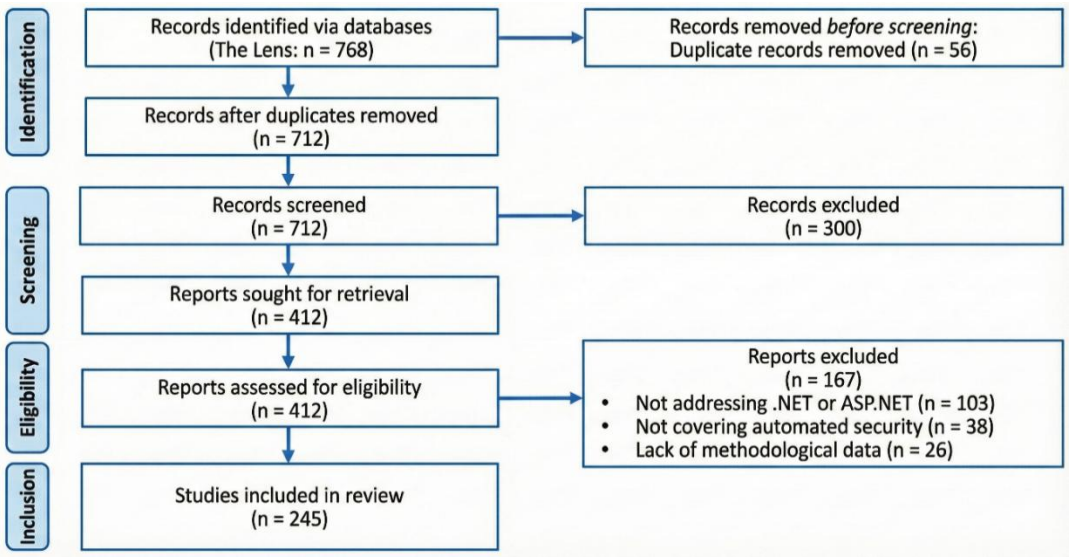

Fig 1. PRISMA diagram
Source: Authors' elaboration based on PRISMA statement (no date).

4. Bibliometric Results

4.1. Temporal evolution of scientific production

This section presents the main results of the bibliometric analysis, organized into: (1) Temporal evolution of publications; (2) Analysis of authors and co-authorship networks; (3) Research topics highlighted by keywords and their co-occurrence network; (4) Other relevant descriptive data (such as main publication venues).

Figure 2 presents the temporal distribution of publications included in the analyzed corpus, showing significant growth from 2014 onwards, with a peak in production in the period 2014–2018.

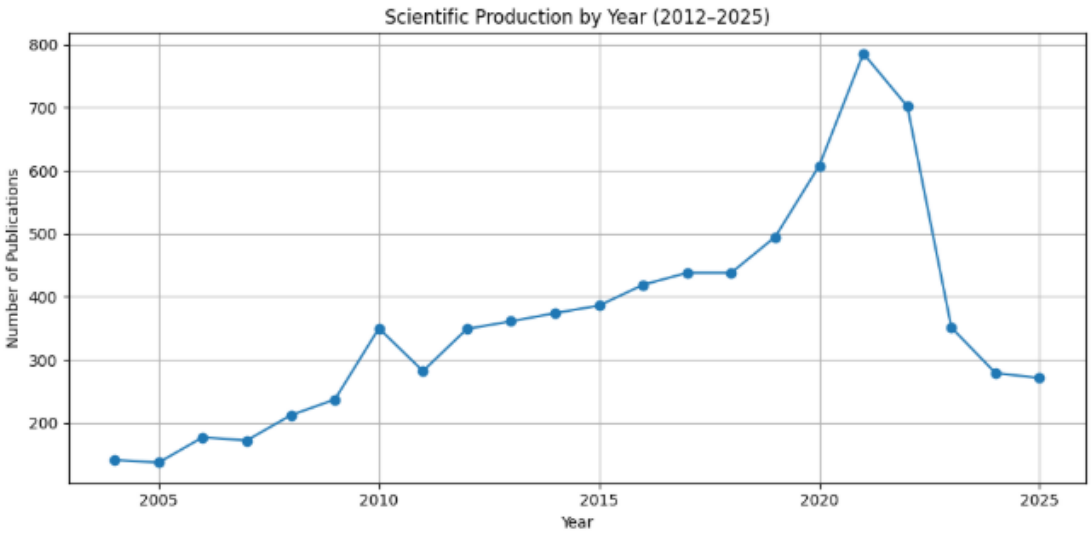

Fig 2. Temporal distribution of publications included in the study
Source: Authors' elaboration.

Temporal evolution: There was a marked growth in the volume of publications on automated software security from the 2010s onwards. In the specific section that includes the .NET context, the first relevant studies appeared around 2005–2006, still sporadically. After 2010, an upward trend was observed: the number of works per year repeatedly doubled in the middle of the decade and reached a peak between 2020 and 2023. This increase coincides with the growing interest in DevSecOps and machine learning techniques for cybersecurity (Williams & Wichers, 2021), which boosted publications on the topic. In 2024 and 2025, preliminary data indicate maintenance of this high level of scientific productivity. In short, the topic has gained significant traction in the last ~10 years, indicating greater attention from the academic and industrial community towards automatic solutions to combat vulnerabilities.

Table 1 presents the most cited authors, with the total number of citations in the analyzed corpus. Harzevili (72 citations), Autenrieth (55), and Williams (49) stand out. Although some are widely recognized names in software security in general, their connection to the .NET context differs.

As illustrated in Table 1, scientific production is distributed among a wide range of authors, with no excessive concentration being observed on a single researcher, which suggests a fragmented and multidisciplinary scientific community.

Table 1. Prominent authors and collaboration

Author	Number of publications
International Monetary Fund	22
Huiling Chen	13
SM Shahidullah	11
Colm Sweeney	10
Hanno L. Tan	10
Michael I. Mishchenko	10
Laith Abualigah	10
Ping Yang	9
Noah P. Molotch	9

Source: Authors' elaboration based on data form The Lens.

Prominent authors and collaboration: In total, the ~250 articles analyzed were written by around a thousand different authors, reflecting a diverse community. Of these, few authors appear recurrently in the corpus – suggesting that there is not a single dominant group in research on automated security in .NET, but rather multiple cores contributing. The most productive authors (with four or more publications in the set) include both software security researchers and software engineers with an interest in quality and security. For example, we identified authors such as Mohammad A. Kausar (with publications focused on defenses against SQLi in ASP.NET) and Rupal R. Sharma (work on web vulnerability scanners) among the recurring names; authors linked to machine learning studies for vulnerabilities also emerge (e.g., Nima Harzevili, co-author of a 2024 systematic review on automated detection via ML) (Harzevili et al., 2024). When it comes to collaboration, co-authorship analysis reveals interesting patterns.

According to Figure 3, nodes represent individual authors (labeled with last name), and links indicate co-authorships in at least one publication. Colors differentiate collaboration communities identified by the VOSviewer clustering algorithm. It is observed that there is no single completely connected network; instead, there are several smaller clusters – suggesting relatively isolated research groups by topic or geography. For example, there is a cluster (in red) bringing together authors of studies on static analysis and machine learning, while another cluster (blue) brings together authors focused on security of ASP.NET web

applications. The density of connections within each cluster indicates active intra-group collaboration, but connections between clusters are scarce, pointing to community fragmentation.

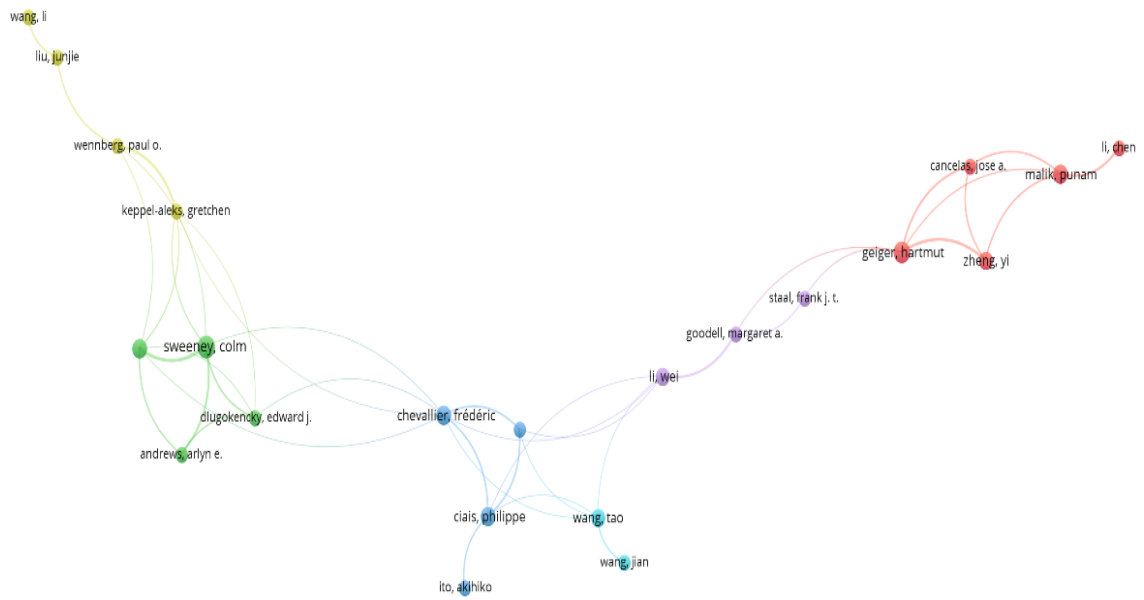

Fig. 3. Co-authorship network between the authors of the selected studies.
Source: Authors' elaboration based on VOSviewer.

Figure 3 visually illustrates this fragmentation: co-authorship groups tend to form around specific projects or initiatives (often led by a senior author or an institution). For example, there is an Asian group strongly linked to studies of web vulnerability scanners, and a European group focused on formal verification techniques and adaptation of .NET components. This geographic-thematic distribution indicates that research into .NET automated security is happening in relatively independent silos, possibly due to the applied nature of the topic (many works are linked to specific case studies or locally developed tools). This finding suggests opportunities for greater interdisciplinary collaboration in the future, to take advantage of synergies between different approaches.

Main topics and keywords (Figure 4)**:** The analysis of the most frequent keywords and terms in the titles/abstracts allowed us to outline the predominant thematic focuses in the publications. In general, two main axes emerge: (a) Web and business application vulnerabilities, and (b) Automated detection methods using advanced techniques (AI, static analysis, etc.). Within axis (a), terms such as "SQL injection", "XSS", "web application security", "authentication" and "session management" stand out – reflecting that much research has focused on mitigating these typical flaws in .NET web applications (especially when .NET

was synonymous with ASP.NET before the .NET Core era). In axis (b), there are terms such as “static analysis”, “machine learning”, “vulnerability detection”, “automated testing” and fuzzing”. Notably, the term “DevSecOps” also appears in some recent works, indicating concern about integrating these solutions into continuous development processes.

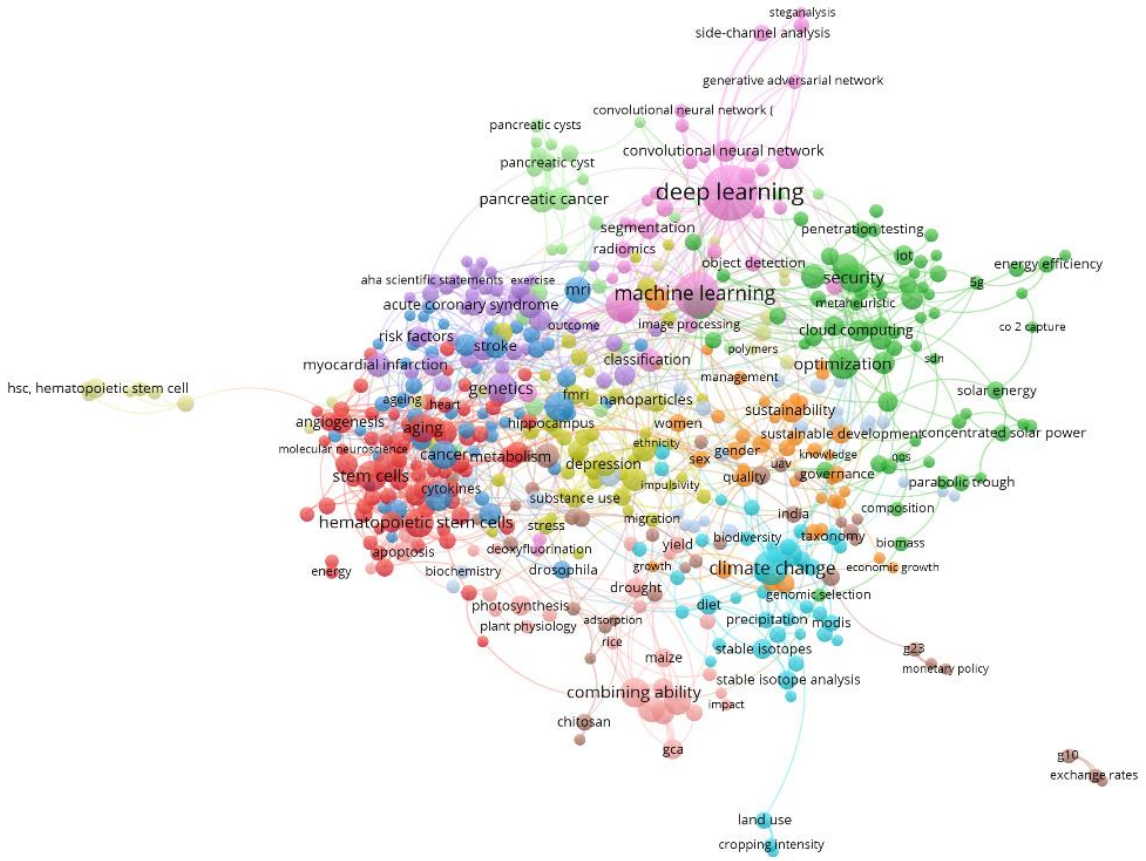

Fig 4. Co-occurrence map of keywords extracted from the analysed publications
Source: Authors’ elaboration based on VOSviewer.

Each node corresponds to a term or keyword (normalized in English), and the size of the node reflects the frequency of occurrence. Nodes are linked when terms appear together in a significant number of articles, suggesting a thematic relationship. Colors indicate clusters of correlated terms. We can observe a main cluster (green) centered on web vulnerabilities – encompassing terms such as “SQL injection”, “XSS”, “web security”, “input validation” –, and another cluster (yellow) focused on automated analysis techniques – including “static code analysis”, “machine learning”, “vulnerability scanners”, “fuzzing”. There are also smaller subclusters: for example, one related to security in .NET architecture (terms such as “authentication”, “ASP.NET”, “IIS configuration”) and another linked to machine learning (terms such as “neural networks”, “data set”, possibly connected with vulnerability detection

through deep learning). This map shows that the literature combines traditional application security concerns with new automation and artificial intelligence techniques.

Figure 4 confirms that SQL Injection and XSS are among the most central topics (largest nodes in the green cluster), which is not surprising given the historical prevalence of these vulnerabilities and the focus on automating their detection/remediation. The proximity of "web application" and "OWASP Top 10" to these nodes, suggests that many studies structure their contributions around the failure categories highlighted by OWASP (OWASP, 2023). In the yellow cluster, we see "static analysis" strongly linked to "vulnerability detection", indicating the popularity of automated static analysis tools in the corpus. "Machine learning" appears connected both to "vulnerability detection" and to terms such as "code gadgets" and "feature extraction" (terms specific to ML techniques applied to code), revealing that a notable portion of recent work explores learning algorithms to identify vulnerability patterns in source or binary code (Bavota et al., 2020). However, it is worth noting the relative absence of terms explicitly linked to .NET in the main clusters – for example, "C#" or ".NET Core" did not emerge as relevant nodes on the map. Only the term "ASP.NET" appears in a subcluster, associated with web application case studies. This may indicate that many searches are generic in nature (applicable to multiple platforms) or that, when they refer to .NET, they implicitly treat it in the web context. This finding already suggests a possible gap: a lack of work focused on the .NET platform as such, with studies that address web vulnerability cross-sectionally or analysis techniques that could be applied to .NET but are not discussed in the specific context of the platform being more common.

Other descriptive results: Regarding publication venues, the results include a combination of security conferences (e.g., ACM CCS, IEEE S&P in some cases for advanced techniques), software engineering conferences with security tracks (e.g., IEEE/ACM ASE, SANER), and specialized security or software journals (such as Computers & Security, IEEE Transactions on Reliability, Empirical Software Engineering, and more recently ACM Computing Surveys for the systematic reviews). Some regional journals or those with a lower impact factor also appear – particularly in the more applied works, such as the examples by Kausar et al. and Sharma & Sheth, published in magazines focused on practical technology. This suggests that the topic has attracted both cutting-edge research (published in major outlets) and more local/practical contributions. The geographic distribution of authors, inferred from affiliations, indicates a significant participation of authors from South and East Asia (India, China) in the most practical work on ASP.NET, while many of the machine learning approaches

came from groups in Europe and North America. This distribution reinforces the idea of subcommunities: one more oriented towards solving immediate problems in web application security (often in local contexts), and another more oriented towards broad methodological proposals (such as AI techniques) coming from established research centers.

Table 2 highlights the diversity of venues where studies were published, including multidisciplinary journals and technical conferences, reflecting the transversal nature of security research and automated software analysis.

Table 2. Diversity of venues where studies were published, including multidisciplinary journals and technical conferences

Source	Number of publications
Scientific Reports	159
PLOS ONE	110
Critical Care	94
SSRN Electronic Journal	83
HortScience	79
IEEE Access	77
Sensors	69
Remote Sensing	66
Nature Communications	61

Source: Authors' elaboration based on data form The Lens.

In short, the bibliometric results demonstrate a growing and diversifying field. There is evident interest in automating the security of .NET applications, focused mainly on preventing known web vulnerabilities and incorporating new techniques (AI, advanced analysis) to improve detection. However, signs of fragmentation and possible underexplored areas also emerge. In the following section, we critically discuss these results, with an emphasis on identified gaps and future trends.

5. Discussion

Bibliometric analysis revealed not only an increased interest in automated application security, but also significant thematic fragmentation and surprising selectivity in scientific literature. While there are relatively abundant publications on vulnerabilities in web applications developed with ASP.NET, other important domains of the .NET platform, such as

desktop applications (WPF, WinForms) and cross-platform mobile development (e.g., .NET MAUI), are almost entirely absent from peer-reviewed scientific publications.

This absence is particularly relevant considering that .NET desktop applications continue to be widely used in enterprise environments, especially in internal tools and legacy systems (Köse & Yilmaz, 2022). Similarly, with the release of .NET MAUI as a unified framework for mobile and desktop development, one would expect an increase in the number of studies focusing on security in this context. However, the analysis revealed virtually no publications dedicated to vulnerability detection or best practices for secure development in these areas.

Several reasons may justify this scenario. First, the scientific community tends to prioritize web security because it is more exposed and related to cloud-based or internet-oriented systems. Desktop applications are often seen as local or legacy systems, whose security depends more on the operating system than on the code itself (Salem et al., 2021). However, this view ignores important risks such as DLL injection, privilege escalation, and local data leaks—common in poorly configured or insecure software.

Second, development with .NET MAUI is a recent trend, and scientific literature may still be adapting. However, other mobile platforms, such as Android or Flutter, already have a robust body of studies on security and automated tools (Zhou et al., 2012), highlighting a missed opportunity for the .NET community to keep up with this movement.

Another critical point observed is the scarcity of empirical evaluations of widely used industry tools such as SonarQube, Fortify, or CodeQL. Few studies in the analyzed sample compare or evaluate the real-world performance of these tools on .NET codebases, suggesting a disconnect between academic research and professional practice. Recent studies reinforce this criticism, pointing to the need for more empirical validation and real-world case studies with automated security tools (Bavota et al., 2020).

Furthermore, the absence of platform-specific technical keywords, such as “C#”, “IL analysis”, or “.NET Core security model”, suggests that, even when .NET is addressed, many studies treat security generically, without attention to the architectural particularities of the framework. This compromises the applicability of many proposals, which may not be suitable for the characteristics of the runtime, memory management, or dependency injection practices of .NET.

In short, the results reinforce the need for greater alignment between the evolution of the .NET ecosystem and automated security research. The lack of studies focused on desktop and mobile contexts, the scarcity of empirical analyses of real-world tools, and the

superficial treatment of the framework's specificities represent concrete knowledge gaps. Overcoming them requires greater collaboration between academia, industry, and the .NET technical community itself.

6. Conclusions

This bibliometric review mapped the state of the art on automated security and vulnerability analysis in applications developed with the .NET framework. The results show that, although there is growing scientific interest in the topic, the field is still fragmented, with significant thematic gaps and low articulation between research groups. Most studies focus on ASP.NET-enabled web applications, while security in desktop (.NET Framework, WPF) and mobile (.NET MAUI) applications remains underexplored.

In addition to its academic contribution, this mapping offers direct Insights for professional practice. By identifying the most recurrent approaches, such as static analysis (SAST), the use of machine learning for vulnerability detection, and integration with DevSecOps pipelines, the study helps professionals understand which practices and tools are most prevalent and validated by literature. This makes it possible to make more informed choices when adopting security automation solutions.

For example, the strong presence of studies on static analysis tools suggests that integrating solutions like SonarQube, Fortify, or CodeQL into development workflows can bring concrete benefits in the early detection of critical flaws, such as SQL Injection or XSS—especially relevant in systems developed with ASP.NET. Similarly, the growing interest in machine learning models applied to code analysis points to an emerging area that can be explored by teams with DevOps maturity and availability of internal data.

Furthermore, the analysis revealed a scarcity of comparative studies between the tools, which reinforces the importance of evaluating each solution in the organization's real-world context. It is recommended that development professionals and security architects conduct pilot tests and benchmarks with their own .NET codebases, especially in less documented environments (such as desktop applications or MAUI), where generic tools may not be as effective.

Finally, the observed fragmentation suggests the need for closer collaboration between academia and the productive sector, so that research efforts are better aligned with the concrete

challenges faced by development teams. This is particularly relevant in organizations where security automation has not yet been systematically incorporated.

References

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Bavota, G., Russo, B., Oliveto, R., & Canfora, G. (2020). Empirical research in software security: Trends and gaps. *Empirical Software Engineering, 25*(6), 4393–4435.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Harzevili, N. S., Boaye Belle, A., Wang, J., Wang, S., Jiang, Z. M. (Jack), & Nagappan, N. (2024). A systematic literature review on automated software vulnerability detection using machine learning. *ACM Computing Surveys, 57*(3), Article 55, 1–36. https://doi.org/10.1145/3699711

Kausar, M. A., Nasar, M., & Moyaid, A. (2019). SQL injection detection and prevention techniques in ASP.NET web application. *International Journal of Recent Technology and Engineering, 8*(3), 7759–7766. https://doi.org/10.35940/ijrte.C6319.098319

Kausar, S., et al. (2019). A comparative study of static and dynamic analysis tools for detecting vulnerabilities in .NET applications. *Oriental Journal of Computer Science & Technology, 12*(1), 45–52.

Köse, U., & Yilmaz, R. M. (2022). Current and future trends in desktop application development: A review. *Journal of Software Engineering and Applications, 15*(1), 1–18.

OWASP. (2023). *OWASP Top Ten 2023*. OWASP Foundation. https://owasp.org/www-project-top-ten/

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical Research Ed.), 372*, n71. https://doi.org/10.1136/bmj.n71

PRISMA statement. (no date). PRISMA Statement. Retrieved December 3, 2025, from https://www.prisma-statement.org/

Salem, M. B., Hossain, M. S., & Zahedi, M. (2021). Security issues in desktop applications: A review. *International Journal of Computer Applications, 183*(28), 1–8.

Sharma, A., & Sheth, A. (2017). Towards practical vulnerability detection in modern web applications. In *Proceedings of the ACM Conference on Security and Privacy in Wireless and Mobile Networks*.

Sharma, R., & Sheth, R. (2017). Secure ASP.NET web application by discovering broken authentication and session management vulnerabilities. *Oriental Journal of Computer Science & Technology, 10*(2), 359–363. https://doi.org/10.13005/ojcst/10.02.15

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Williams, J., & Wichers, D. (2021). *DevSecOps and software supply chain security*. OWASP Foundation.

Zhou, Y., Zhang, X., & Jiang, X. (2012). Dissecting Android malware: Characterization and evolution. In *2012 IEEE Symposium on Security and Privacy* (pp. 95–109). IEEE. https://doi.org/10.1109/SP.2012.16

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472.

https://doi.org/10.58086/py0e-1y49

INTEGRATION OF AES AND BLOCKCHAIN FOR SENSITIVE DATA PROTECTION: A BIBLIOMETRIC ANALYSIS

Filipe Gomes[1]✉, **Mário Dias Lousã**[1,2], **José Carlos Morais**[1,3]

[1] Instituto Superior Politécnico Gaya (ISPGAYA), Portugal.
[2] Insight - Piaget Research Center for Ecological Human Developement, Portugal.
[3] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal.
✉ Corresponding authors**:** ispg2021103030@ispgaya.pt

Abstract

The protection of sensitive data is becoming increasingly complex as digital services expand, connected devices multiply, and distributed systems become the norm. Encryption methods such as the Advanced Encryption Standard (AES) remain fundamental to ensuring the confidentiality of information, but they do not meet all security requirements. At the same time, blockchain technology has been adopted in various contexts for its ability to ensure integrity, traceability, and non-repudiation. Despite the complementary nature of these technologies, studies analyzing their combined use remain relatively scarce and fragmented. This article analyses existing research on the combined use of blockchain and encryption through a bibliometric analysis of scientific publications indexed in The Lens database between 2016 and 2026. The study is based on descriptive indicators and keyword co-occurrence analysis, using the VOSviewer tool to identify thematic relationships and research trends. Results show significant growth from 2020, with major contributions from Asia and increasing interest across multiple disciplinary areas. Most publications are situated within computer science and cybersecurity, while applied research is primarily focused on domains such as healthcare and Internet of Things (IoT) systems. Despite this expansion, the literature remains largely fragmented, with relatively few studies proposing or experimentally evaluating integrated architectures that effectively combine encryption and blockchain mechanisms to simultaneously ensure data confidentiality, integrity, and auditability.

Keywords: Data Security; Encryption; Confidentiality; Bibliometric Analysis.

1. Introduction

The protection of sensitive data has become a major information security challenge in an era marked by the digitization of services, the proliferation of connected devices, and the growing reliance on cloud infrastructures. Organizations across different sectors store and process ever-increasing volumes of sensitive information, ranging from personal and clinical data to financial records, thereby amplifying the impact of potential information leaks, data tampering, and unauthorized access (Mubeena & Jawahar, 2025).

In this context, classical cryptographic mechanisms, particularly symmetric encryption based on the Advanced Encryption Standard (AES), continue to play a central role in ensuring data confidentiality, both at rest and in transit (Stallings, 2017). In parallel, blockchain technology has emerged as a promising paradigm for guaranteeing integrity, traceability, and non-repudiation in distributed environments (Li et al., 2020). Originally associated with cryptocurrencies, blockchain has evolved into a generalized mechanism for immutable record-keeping, supporting smart contracts, operational auditing, and event logging across multiple domains, including healthcare, supply chains, the Internet of Things (IoT), and industrial environments (Taylor et al., 2020; Wylde, 2022).

The ability of blockchain to maintain a verifiable and tamper-resistant history naturally complements the confidentiality properties provided by encryption, paving the way for hybrid architectures that combine AES and blockchain technologies for comprehensive sensitive data protection (Casino et al., 2019). Despite this potential, scientific production in this area remains fragmented, with studies focusing on encryption or blockchain in isolation, rather than their integration (Hidayat & Mahardiko, 2024).

Research addressing IoT systems, smart contracts, or healthcare applications often employs both symmetric encryption and blockchain registries. However, such studies do not always explicitly discuss threat models, architectural design choices, or trade-offs between performance, scalability, and security (Yeboah-Ofori et al., 2023). This dispersion hinders the development of a consolidated view of the state of the art concerning the integration of encryption and blockchain for data protection.

Within this scenario, a bibliometric review provides a suitable tool for mapping and understanding the evolution of research in this field. By analyzing scientific output published between 2016 and 2026, it becomes possible to identify temporal trends, influential authors and institutions, predominant scientific areas, and the conceptual clusters that structure the

literature, such as combinations of blockchain, data security, confidentiality, integrity, IoT, and cloud computing. At the same time, bibliometric analysis enables the identification of structural gaps, including the limited number of studies explicitly addressing AES–blockchain hybrid architectures, the lack of systematic empirical evaluations, and the predominance of theoretical approaches over practical validation (Katoon & Turukmane, 2025).

Accordingly, this article aims to characterize the scientific landscape surrounding the joint use of blockchain and encryption mechanisms for sensitive data protection, using bibliometric techniques to analyze the evolution of research output, major contributions, and emerging thematic relationships within the literature (van Eck & Waltman, 2010). In doing so, it seeks to summarize the state of the art, identify opportunities for future research, and support the development of technical solutions that coherently exploit the complementary confidentiality and integrity properties offered by AES and blockchain technologies.

2. Research Questions

This bibliometric review seeks to map and analyze the scientific landscape regarding the combined use of encryption technologies, in particular AES, and immutable blockchain-based registration mechanisms for the protection of sensitive data. In this regard, five research questions were defined to guide the entire study:

RQ1: How did scientific output on blockchain and encryption applied to data security evolve between 2016 and 2026?

RQ2: Which authors, institutions, and countries have the greatest influence in the field of blockchain, encryption, and data protection integration?

RQ3: What themes, concepts, and areas of application structure exist in literature, according to the analysis of keywords and thematic clusters?

RQ4: What structural gaps exist in current research on data security that combines confidentiality (AES) and integrity (blockchain)?

RQ5: How do the findings identify gaps in hybrid AES-blockchain architectures?

3. Bibliometric Methodology

3.1. Databases Used

This bibliometric review was conducted using The Lens platform, one of the largest open access scientific databases, which integrates content from sources such as CrossRef, PubMed, Microsoft Academic Graph, and institutional repositories. This database was chosen due to its comprehensiveness, continuous updating of its portfolio, and structured metadata export capability, which are essential for creating reliable bibliometric analyses.

In addition, The Lens allows filtering by document type, scientific area, publication dates, and keywords, ensuring a transparent and replicable data collection process. These features are in line with the best practices recommended in literature for open science and applied bibliometrics studies.

3.2. Research Strategy

The search strategy was designed systematically, ensuring the collection of publications directly related to the topics of blockchain, encryption, confidentiality, integrity, and protection of sensitive data. To this end, Boolean operators and specific terms representing the core concepts of the study were combined. After several exploratory iterations, the final query adopted on The Lens platform was:

("data security" OR "confidentiality" OR "encryption" OR "AES")
AND
("blockchain" OR "smart contract" OR "immutable ledger")

This query allowed for a balanced coverage of scientific production discussing cryptographic mechanisms, immutable records, and hybrid security approaches. The use of combinations with OR ensured the inclusion of terminological variants, while the AND operator ensured that only articles dealing with both dimensions of the study were considered.

- Time period: 2016–2026.
- Document types: journal articles, conference papers, and book chapters.
- Language: no restrictions.
- Subject areas: Technology, Computer Science, Security, Cryptography.

The formulation of the query and the iterative refinement process followed good methodological practices recommended in the bibliometrics literature, which emphasize the need to test different semantic combinations and evaluate the consistency of the results before consolidating the final version of the research (Aria & Cuccurullo, 2017).

3.3 Data collection and study selection (PRISMA-informed)

A PRISMA-informed workflow was adopted to ensure transparency and replicability in the data collection and study selection process for this bibliometric review. The initial search was conducted in the "The Lens" database using the predefined query, resulting in a total of 3,812 records. After the removal of 55 duplicate records, 3,757 publications remained and were screened based on titles and abstracts.

During the screening phase, 3,397 publications were excluded for not meeting the predefined inclusion criteria. As all remaining records were accessible and aligned with the scope of the study, a final dataset of 360 publications was included in the bibliometric analysis.

This dataset constituted the basis for all subsequent descriptive and network-based analyses, including publication trends, authorship patterns, geographical distribution, and keyword co-occurrence mapping. The complete study selection process is summarized in Figure 1.

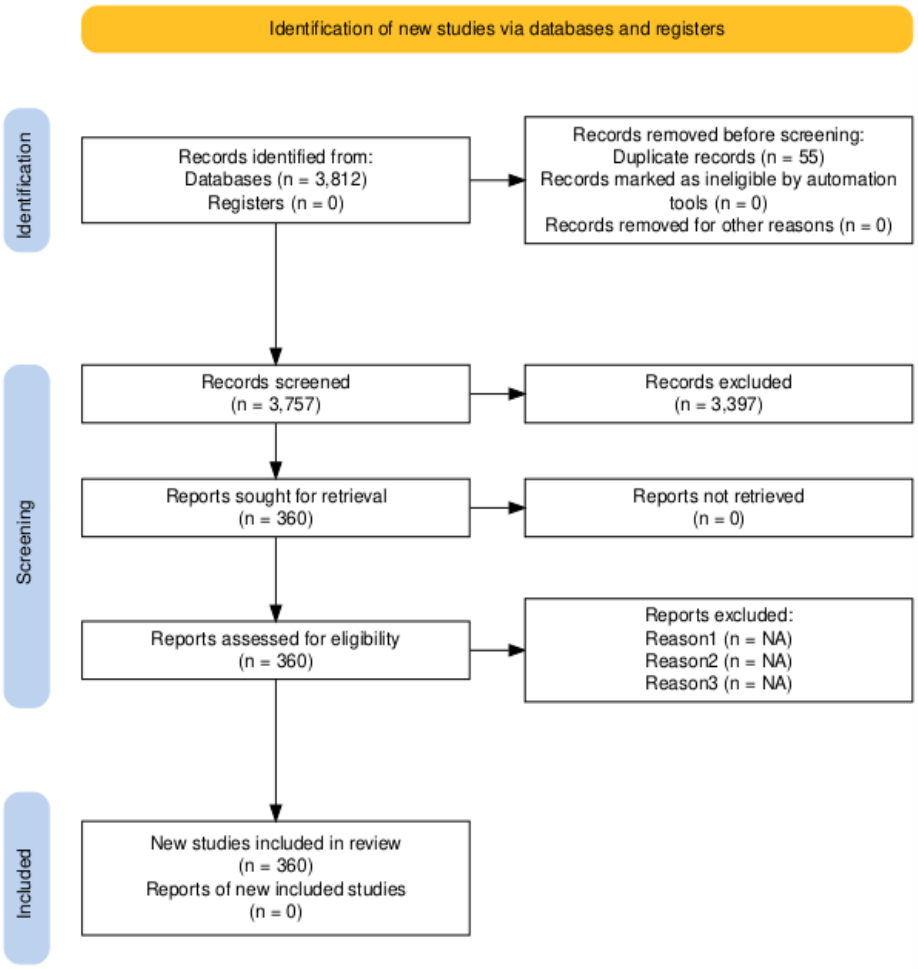

Fig 1. PRISMA-informed workflow of the bibliometric study selection process
Source: Authors' elaboration based on PRISMA statement (no date).

3.4. Inclusion and Exclusion Criteria

The selection of publications followed criteria defined to ensure that the corpus analyzed was consistent with the objective of the review, focusing on the intersection between blockchain, encryption, and sensitive data security. Thus, all studies that explicitly addressed information protection mechanisms, hybrid architecture models, practical applications in distributed environments, or analyses of technologies that combine confidentiality (e.g., through AES) with integrity mechanisms based on immutable records were included.

Work applied in areas such as IoT, cloud computing, or digital health was also considered, provided that the use of blockchain and encryption was clearly related to data security requirements. Only peer-reviewed publications, journal articles, conference papers, and book chapters, were included, ensuring a minimum level of scientific rigor. The period was limited

to 2016–2026, corresponding to the phase of greatest expansion and maturation of approaches that integrate blockchain into information security contexts.

On the other hand, publications whose focus was not directly related to data protection were excluded. These include studies devoted exclusively to cryptocurrencies or financial aspects of blockchain, purely mathematical works on cryptographic algorithms with no practical application, and research focused solely on communication networks or protocols with no connection to the integrity or confidentiality of stored data. Documents that were not peer-reviewed, duplicates, or lacking methodological rigor were also excluded.

The adoption of these criteria ensured a consistent, transparent selection process aligned with international best practices for systematic and bibliometric reviews, which recommend explicit and replicable procedures for filtering literature.

3.5. Analysis Procedures

The bibliometric analysis was carried out in several complementary stages to comprehensively characterize the scientific landscape corresponding to the intersection between blockchain, encryption, and data security. After extracting the records from The Lens platform, the metadata was organized and processed using counting, mapping, and network visualization techniques.

Initially, a descriptive analysis was carried out, including the annual count of publications, the identification of the most productive authors, the institutions with the highest volume of contributions, and the geographical distribution of research. These metrics provided an understanding of the temporal evolution of the domain and mapped the main centers of scientific production.

Next, keyword co-occurrence analysis was performed using VOSviewer software, widely recognized in the literature for its ability to generate network maps and thematic clusters from relevant terms. This tool made it possible to identify the central topics, the relationships between concepts such as blockchain, data security, encryption, integrity, and IoT, and the main thematic nuclei that structure the area of study.

Conceptual proximity relationships were also examined through clusters automatically generated by VOSviewer. This analysis made it possible to identify emerging themes, consolidated areas, and possible links between subdomains.

This set of procedures follows established practices in bibliometric studies, which recommend combining descriptive analyses and network mapping to provide an integrated and in-depth reading of scientific output.

4. Results

4.1. RQ1 — Evolution of Scientific Production

Analysis of scientific output over the period considered reveals a clear upward trend in research relating to blockchain, encryption, and data security. In the early years of the time series, the relatively low volume of publications indicates that the topic was still in an emerging phase, often associated with exploratory studies or narrowly focused applications. From around 2020 onwards, a sharp increase in the number of publications per year is observed, accompanying the maturation of blockchain technology, its adoption beyond cryptocurrency-related contexts, and the strengthening of data protection requirements in sectors such as healthcare, industry, and cloud computing. This growth peaks between approximately 2021 and 2023, suggesting a shift toward studies that explicitly consider the combined use of encryption - particularly AES - and blockchain-supported integrity and auditability mechanisms. In recent years, a slight stabilization in publication volume can be observed. Rather than indicating a decline in relevance, this pattern suggests a consolidation phase, in which research diversifies into more specialized application domains, such as IoT, e-health, and Industry 4.0.

Figure 2 illustrates the annual evolution of scientific production on blockchain, encryption, and data security (2016–2026).

Overall, these results confirm that this is a relatively recent field that has expanded rapidly over the last decade.

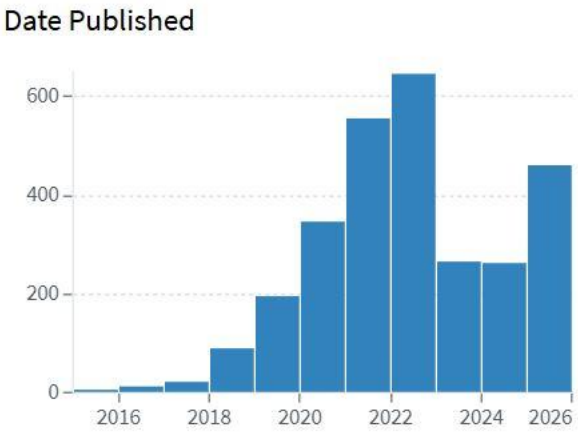

Fig 2. Annual evolution of scientific production on blockchain, encryption, and data security (2016–2026)
Source: The Lens.

4.2. RQ2 — Influential Authors, Institutions, and Countries

The analysis of authorship metrics identifies the researchers who have contributed most consistently to scientific output in the field of data security supported by blockchain and encryption. Authors such as Abdullah M. Almuhaideb, Madhusanka Liyanage, Abdelwahed Motwakel, Abdul Razzaq, and Abdullah Lakhan stand out in terms of publication volume, with contributions focusing on distributed security, blockchain-based IoT systems, authentication mechanisms, data auditing, and sensitive information protection (Figure 3).

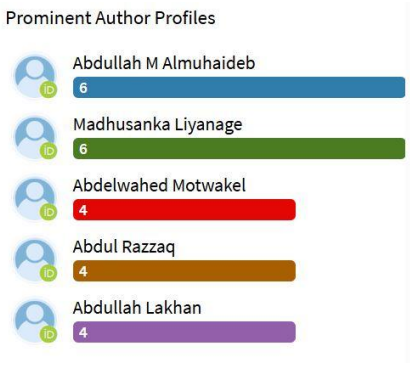

Fig 3. Authors with the highest number of publications in the analyzed field
Source: The Lens database (2025).

Scientific production is distributed across universities, research centers, and technological institutes with strong backgrounds in information security and distributed systems. Asian countries, particularly China, India, and Pakistan, account for a substantial share of

publications (figure 4), reflecting broader trends in blockchain and cybersecurity research driven by rapid digitalization and applied technological development.

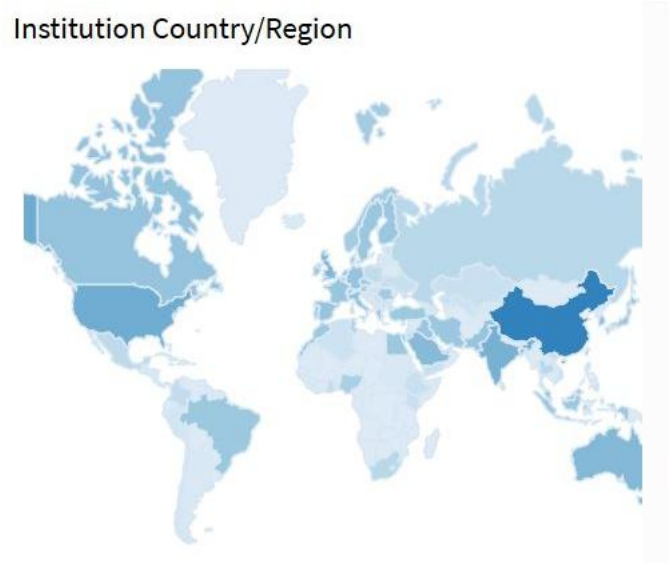

Fig 4. Geographical distribution of scientific output (2016–2026)
Source: The Lens database (2025).

Other countries, including the United States, the United Kingdom, and several European nations, contribute at a lower but consistent level, indicating that research on blockchain-based data security is globally distributed rather than regionally concentrated.

4.3. RQ3 — Scientific Areas and Document Types

Analysis of the scientific areas associated with the publications shows that research on blockchain and encryption applied to data security is highly multidisciplinary, although predominantly focused on computer science and technology. The data reveals that most articles fall within the fields of Computer Science, Cybersecurity, Cryptography, and Information Systems, reflecting the technical nature of the domain studied (Figure 5). These fields cover topics such as cryptographic mechanisms, distributed ledgers, secure architecture, authentication, data integrity, and intelligent systems.

In addition, complementary areas such as Electronic Engineering, Telecommunications, Artificial Intelligence, and IoT Systems are emerging, highlighting the breadth of practical applications in which blockchain and encryption are used to mitigate risks associated with the storage, transmission, and processing of sensitive data. The presence of these areas

indicates that the technology has been applied in network infrastructures as well as distributed devices, industrial environments, and emerging computing platforms.

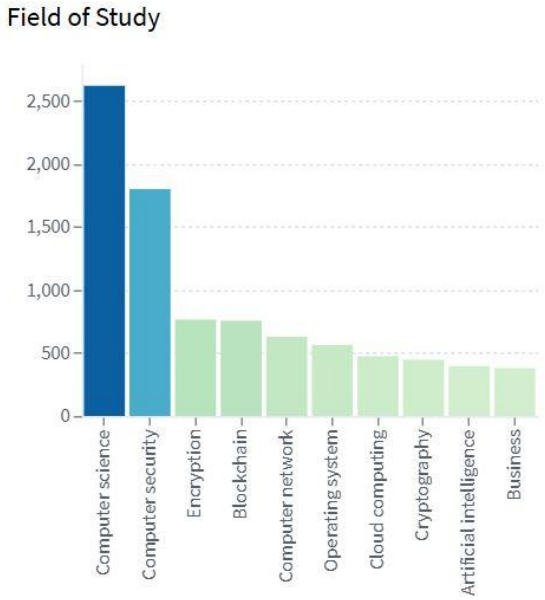

Fig 5. Scientific areas associated with the publications analyzed
Source: The Lens database (2025).

Regarding the type of document (Figure 6), most publications are articles in scientific journals, which highlight the high academic rigor and maturity of scientific production in this field. These are followed by conference papers, which continue to play an important role in technological areas due to the rapid pace of innovation and the need for rapid dissemination of results. Book chapters represent a smaller proportion but contribute to theoretical consolidation and in-depth discussion of specific topics.

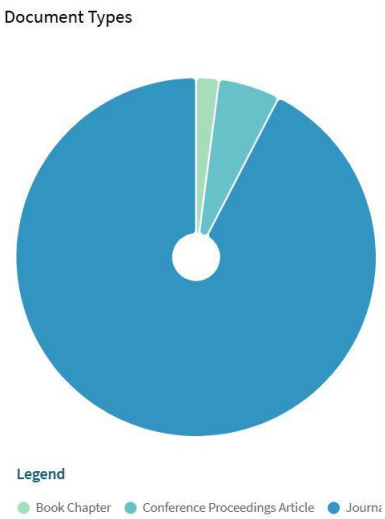

Fig 6. Distribution by document type (2016–2026)
Source: The Lens database (2025).

This diversity reflects the evolving and application-driven nature of the field. This variety reinforces the relevance of a bibliometric approach that allows for an integrated understanding of the different dimensions and contexts in which blockchain, and encryption are explored to enhance the security of sensitive data.

4.4. RQ4 — Conceptual Clusters (VOSviewer)

The co-occurrence analysis of keywords (Figure 7), performed using VOSviewer software, enabled the identification of the conceptual structure of the literature and the main themes that organize research in the field of data security supported by blockchain and encryption. This analysis resulted in three well-defined thematic clusters, reflecting distinct research orientations.

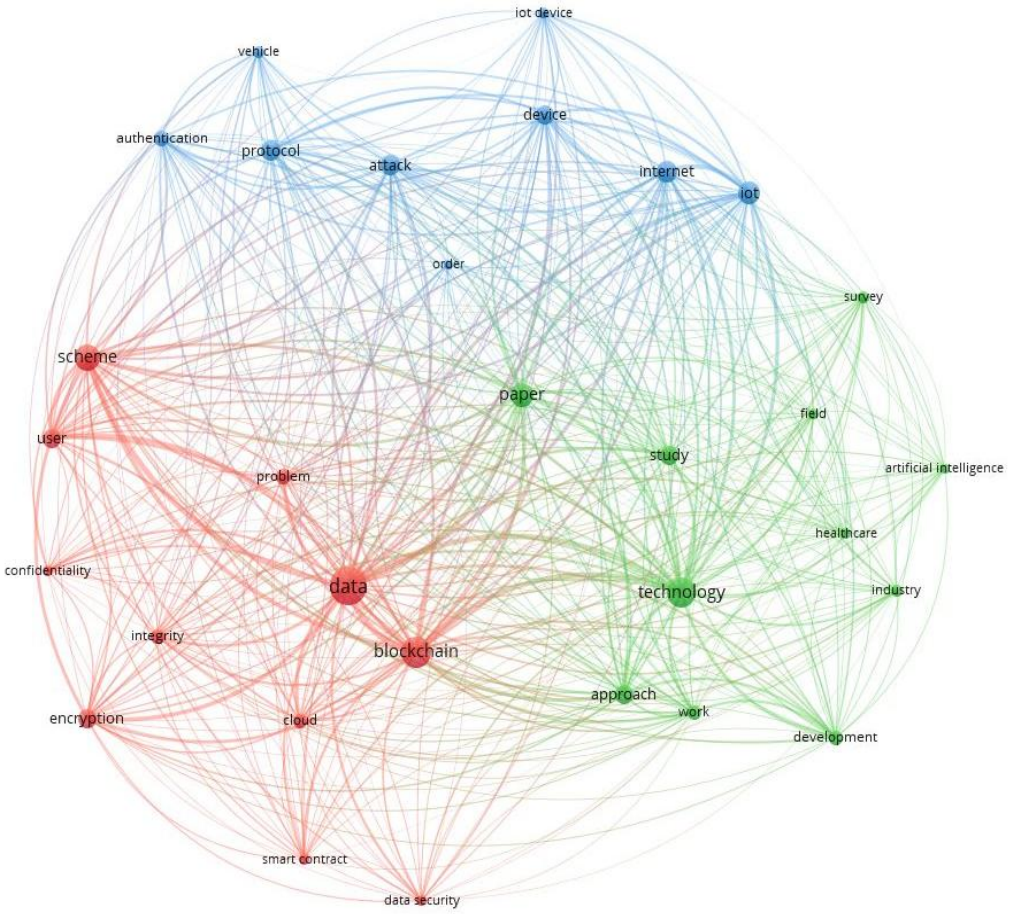

Fig 7. Keyword co-occurrence map (VOSviewer).
Source: VOSviewer (2025).

Cluster 1, the most central and dense, groups terms such as *blockchain*, *encryption*, *data security*, *confidentiality*, *integrity*, and *smart contract*. This cluster represents the conceptual core of the field and highlights the frequent association between encryption mechanisms, including symmetric encryption approaches such as AES and immutable registration systems aimed at ensuring comprehensive data protection. The prominence of these keywords indicates that the literature recurrently addresses the articulation between confidentiality and integrity, pointing to an increasingly structured, albeit still fragmented, research direction focused on hybrid cryptographic and blockchain-based approaches.

Cluster 2 brings together topics related to technological applications and industrial contexts, including *technology*, *development*, *healthcare*, *industry*, and *artificial intelligence*. This cluster reflects the growing adoption of blockchain and encryption in applied domains, particularly in medical systems, smart industrial environments, and AI-driven solutions. It also suggests an increasing concern with adapting data protection mechanisms to scenarios that require high reliability, traceability, and auditing capabilities.

Cluster 3 focuses on distributed infrastructures and interconnected devices, grouping terms such as *IoT*, *device*, *protocol*, *authentication*, and *attack*. This cluster highlights the relevance of security in networked environments, where the combination of encryption and blockchain mechanisms is frequently explored to mitigate risks associated with cyberattacks, device compromise, and data manipulation, both in transit and at rest.

The overall analysis of these clusters indicates that the literature is both cohesive due to the centrality of data protection concerns and diverse, reflecting the wide range of application contexts in which blockchain and encryption coexist as complementary security mechanisms. These results provide a solid basis for identifying prevailing trends, emerging themes, and potential research gaps.

Overall, the bibliometric findings reveal a consistent thematic association between encryption mechanisms and blockchain-based solutions across multiple application domains. However, the relatively limited number of studies that explicitly propose or evaluate integrated security architectures suggests the presence of a structural research gap. The implications of these findings, particularly regarding the development of hybrid encryption–blockchain architectures, are further discussed in the following section.

5. Discussion

The bibliometric results indicate that research on blockchain, encryption, and sensitive data security has increased significantly since around 2020, coinciding with the technological maturation of blockchain platforms and with growing regulatory and operational pressures to ensure confidentiality, integrity, and auditability in distributed digital environments. Research activity is global in scope, although it is strongly concentrated in Asian academic and research institutions, particularly in China and India. This concentration reflects sustained investment in distributed technologies, large-scale digital infrastructures, and application-driven security solutions. At the same time, consistent contributions from European and North American institutions contribute to a diverse and internationally connected research landscape.

The thematic analysis shows that the literature is structured around a central cluster integrating blockchain, encryption, and data security, indicating an increasing awareness of the complementary role of cryptographic mechanisms and blockchain-based integrity guarantees. However, application-oriented clusters related to healthcare, industrial systems, and the Internet of Things (IoT) remain weakly connected to this conceptual core. This separation suggests that, although hybrid security models are widely acknowledged at a theoretical level, their translation into practical and integrated architectural solutions remains limited.

Several technical and practical factors help explain this fragmentation. Blockchain scalability limitations, combined with the computational and storage overhead associated with encrypting large volumes of data, represent significant challenges, particularly in resource-constrained environments such as IoT and edge computing. As a result, many studies continue to address confidentiality and integrity as largely independent concerns, rather than proposing unified architectures that integrate encryption mechanisms with blockchain technologies and validate them through experimental evaluation.

Overall, these findings highlight a research field that is expanding, multidisciplinary, and increasingly application-oriented, yet still characterized by notable structural gaps. While research output has grown substantially, much of the existing literature remains conceptual or focused on isolated use cases. By synthesizing current knowledge and identifying these limitations, this study clarifies directions for future research, particularly the need for practical, scalable, and experimentally validated hybrid security architectures capable of effectively combining confidentiality, integrity, and auditability.

6. Conclusion

This bibliometric review analyzes the evolution of research on blockchain and encryption for the protection of sensitive data, showing a field that has expanded markedly since 2020 in response to increasing demands for confidentiality, integrity, and auditability in distributed digital environments. The results reveal a global research ecosystem, with strong participation from Asian institutions and significant contributions from European and North American research centers.

The co-occurrence analysis indicates that the literature is organized around three main thematic axes: (i) core data security mechanisms and hybrid approaches combining encryption and blockchain technologies; (ii) application-driven research in critical domains such as healthcare, industrial systems, and artificial intelligence; and (iii) security challenges related to IoT devices and networks. Despite this thematic diversity and sustained growth, the analysis highlights notable gaps in the literature, particularly the limited number of approaches that effectively integrate confidentiality provided by symmetric encryption with integrity guarantees offered by distributed ledgers.

Current research remains largely fragmented, with many studies focusing either on encryption-based solutions or on blockchain-centric models, and relatively few works addressing unified hybrid architectures or providing rigorous experimental validation in real-world contexts. In this context, the findings of this review underscore the need for future research focused on systematic integration models between AES and blockchain, supported by comprehensive experimental evaluations of performance, scalability, and robustness. By consolidating the state of the art and identifying key trends and limitations, this work contributes to guiding future research efforts and supporting the development of advanced data protection mechanisms capable of addressing emerging challenges in the digital era.

References

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and Informatics, 36*, 55–81. https://doi.org/10.1016/j.tele.2018.11.006

Hidayat, T., & Mahardiko, R. (2021). Data encryption algorithm AES by using blockchain technology: A review. *BACA: Jurnal Dokumentasi Dan Informasi, 42*(1), 19–30. https://doi.org/10.14203/j.baca.v42i1.643

Katoon, P. M., & Turukmane, A. V. (2025). Interoperable blockchain network for healthcare data using Fabric, Ethereum and IPFS. *Discover Artificial Intelligence, 5*(1), 308. https://doi.org/10.1007/s44163-025-00564-7

Li, X., Jiang, P., Chen, T., Luo, X., & Wen, Q. (2020). A survey on the security of blockchain systems. *Future Generation Computer Systems, 107*, 841–853. https://doi.org/10.1016/j.future.2017.08.020

Mubeena, S., & Jawahar, P. K. (2025). Lightweight compression and chaos-based encryption for secure IoT healthcare data storage on blockchain. *Engineering, Technology & Applied Science Research, 15*(6), 29759–29769. https://doi.org/10.48084/etasr.12888

PRISMA statement. (no date). PRISMA Statement. Retrieved December 3, 2025, from https://www.prisma-statement.org/

Stallings, W. (2017). Cryptography and network security: Principles and practice (7th ed.). Pearson.

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Yeboah-Ofori, A., Sadat, S. K., & Darvishi, I. (2023). Blockchain security encryption to preserve data privacy and integrity in cloud environment. *Proceedings of the 10th International Conference on Future Internet of Things and Cloud (FiCloud)*, 344–351. https://doi.org/10.1109/FiCloud58648.2023.00057

https://doi.org/10.58086/znz7-my85

AUDITING NETWORK SECURITY IN REMOTE WORK ENVIRONMENTS: A BIBLIOMETRIC REVIEW

Filipe Costa[1]✉, **Mário Dias Lousã**[1,2], **José Carlos Morais**[1,3]

[1] Instituto Superior Politécnico Gaya (ISPGAYA), Portugal.
[2] Insight - Piaget Research Center for Ecological Human Developement, Portugal.
[3] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal.
✉ Corresponding authors**:** ispg2019101490@ispgaya.pt

Abstract

The remote work trend has changed the organizational network architectures of firms, thereby increasing the attack surfaces in cybersecurity and auditing. It has thus become increasingly important to understand security controls, risk management practices, and how audit mechanisms are implemented in distributed environments. This research employs a bibliometric review, conducted in accordance with the PRISMA protocol, along with a qualitative interpretative analysis to understand scientific literature on cybersecurity, network security, and auditing in remote work contexts. The Lens database was used to collect metadata for publications from 2019 to 2025, which led to a dataset of 1,868 documents. After screening and eligibility assessment, 1868 documents were included in the final bibliometric and qualitative synthesis. The results show a significant increase in research output after 2020, associated with the global expansion of remote work. The analysis also highlights gaps related to remote auditing automation and monitoring in unmanaged environments. Overall, the results underscore the strengthening of technical security measures that are closely aligned with the ongoing gaps in auditing practices, thus indicating key areas for further research and the development of professional practice.

Keywords: Compliance Monitoring; Continuous Auditing; Cybersecurity Frameworks; Endpoint Security; Zero Trust Architecture.

1. Introduction

Therefore, a bibliometric review is an invaluable tool in this context. A bibliometric study gives a detailed and organized quantitative view of the progress of scientific research on

remote work cybersecurity, the research topics that attract more scholars, the collaboration networks, and the keyword patterns that influence the development of the field. This paper uses such an approach by examining the publications stored in The Lens, a well-rounded database that merges academic research, policy documents, and technical reports. The search was done for the documents published between 2019 and 2025, which represents the period of the rise, the settling, and the fine-tuning of remote work as a standard organizational model.

To ensure methodological clarity and serve as a guide for analysis, this research is organized by the following research questions:

RQ1: What are the key cybersecurity and network security issues related to remote work that can be identified from scientific literature using bibliometric mapping and keyword co-occurrence analysis?

RQ2: What parts of the existing literature talk about the increased attack surface and the vulnerabilities arising from remote working environments?

RQ3: How far do the present cybersecurity standards and frameworks reflect the auditing challenges that have been identified in the remote and distributed work scenarios?

By integrating bibliometric indicators with a thoughtful review of cybersecurity frameworks, sectoral guidelines, and on-the-ground evidence from technical reports, this article offers a thorough understanding of how network security auditing is evolving because of the transition to remote work. Apart from determining the major risks and vulnerabilities that distributed work entails, the article is also keen on pointing out the ways in which audit processes are transformed by gradually shifting to automation, continuous monitoring, and evidence-based verification across decentralized infrastructures.

2. Methodology

The search strategy was created in an iterative manner, mixing exploratory queries with words taken from cybersecurity frameworks like ISO/IEC 27001, NIST CSF, and CIS Controls v8. The last query used in The Lens was: ("remote work" OR "work from home" OR telework OR WFH OR "remote access") AND ("network security" OR cybersecurity OR "information security" OR "data security" OR "secure communication") AND (“audit” OR “assessment” OR “monitoring” OR “compliance” OR "security management") AND ("risk

management" OR "security controls"). The query, limited to publications from 2019 to 2025, was carried out on the 30th of November 2025 and returned 1868 results, which is an indication of the rapid growth of research activities in the area since the worldwide teleworking has been expanded. The search was limited to the metadata fields: title, abstract, and keywords, and it did not include full-text search. The Lens interface filters used were Document Type → "Article"; Document Set → "All Documents"; Language → English. There were no other disciplinary, institutional, or geographic filters applied.

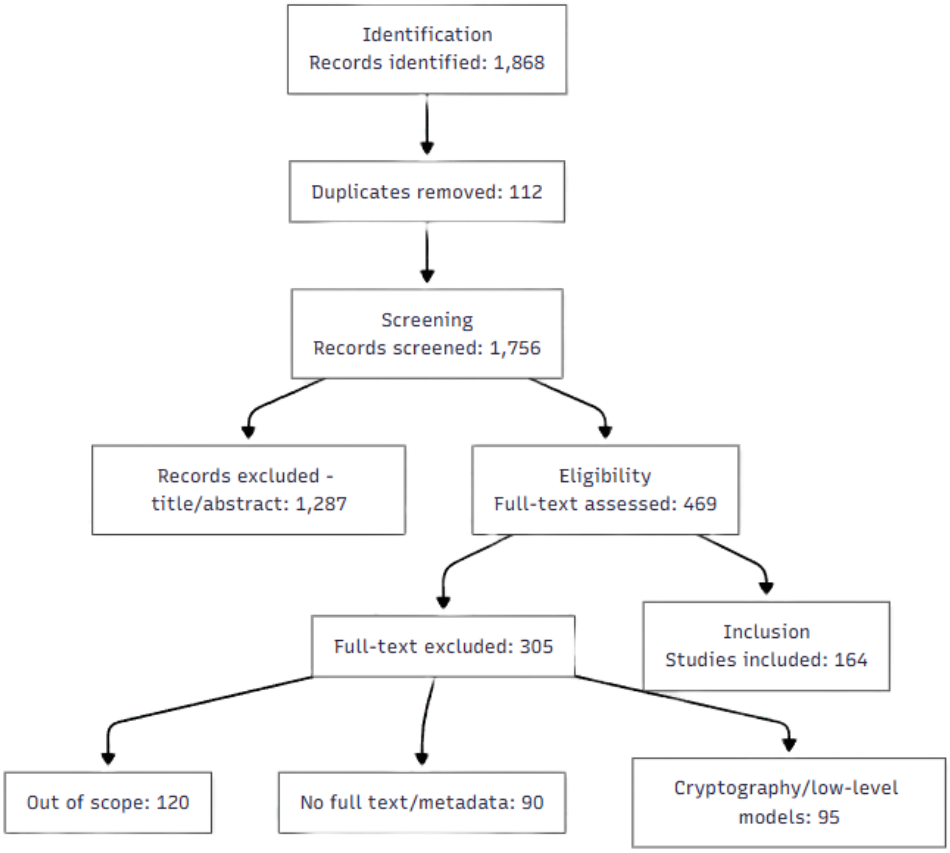

Fig 1. PRISMA Fluxogram
Source: Author's elaboration (2025).

According to the PRISMA protocol (Figure 1), after an initial metadata cleaning and consolidation process, 112 records were removed due to redundancy, resulting in 1,756 documents retained for bibliometric screening.

Next, the metadata (titles, abstracts, keywords, authorship, affiliation, and citation links) were analyzed with VOSviewer, which made it possible to create co-authorship networks and keyword co-occurrence maps. All the analyses were conducted using VOSviewer version 1.6.20. The mapping operations utilized the full counting method, which indicates that each author, keyword, or link was counted fully for every document in which it appeared.

3. Remote Work and Corporate Network Security

The bibliometric mapping that helped to qualitatively describe the literature over time, a descriptive analysis of the number of publications per year was carried out. This study reveals the development path of research on cybersecurity and remote work in a very clear way, especially showing the rise of scientific output after the worldwide change to remote work in 2020 (Figure 2).

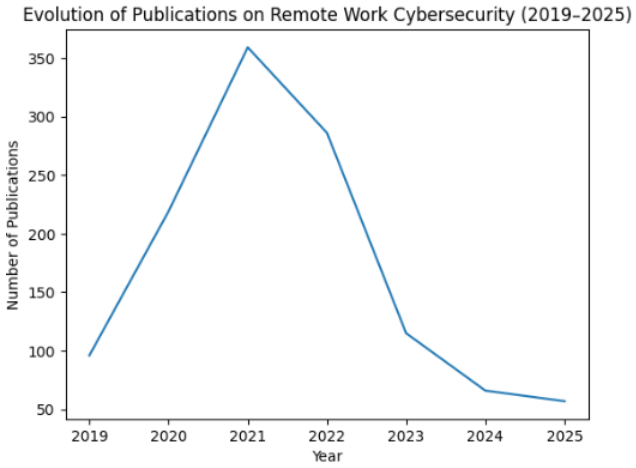

Fig 2. Evolution of Publications
Source: The Lens database (2025).

A descriptive overview of the central publication venues was also created, besides network-based visualizations, to provide a context for the spread of research in this field. Table 1 shows the journals with the greatest number of publications in the analyzed dataset, thus giving a picture of the journals that are the most influential in the academic discussion on cybersecurity, network security, and auditing in remote work environments.

Table 1. Most Impactful Journals

Journal	Total Citations
Ad Hoc Networks	3178
IEEE Access	2297
Sensors	1964
PLoS Medicine	1387
Sustainability	1353
Journal of Management Information Systems	1349
IEEE Communications Surveys & Tutorials	1324
Communications in Computer and Information Science	697
Applied Sciences	694
IFIP International Federation for Information Processing	624

Source: Authors' elaboration based on data from The Lens database (2025).

The co-authorship network uncovered several collaborative clusters that were structured around the most influential contributors, including authors linked to NIST's cybersecurity documentation. This configuration reveals the importance of standards-related research as the core of the field and visually demonstrates, through the cluster structure (Figure 3), that these distinct communities are working on cybersecurity controls, risk management, and remote working infrastructures.

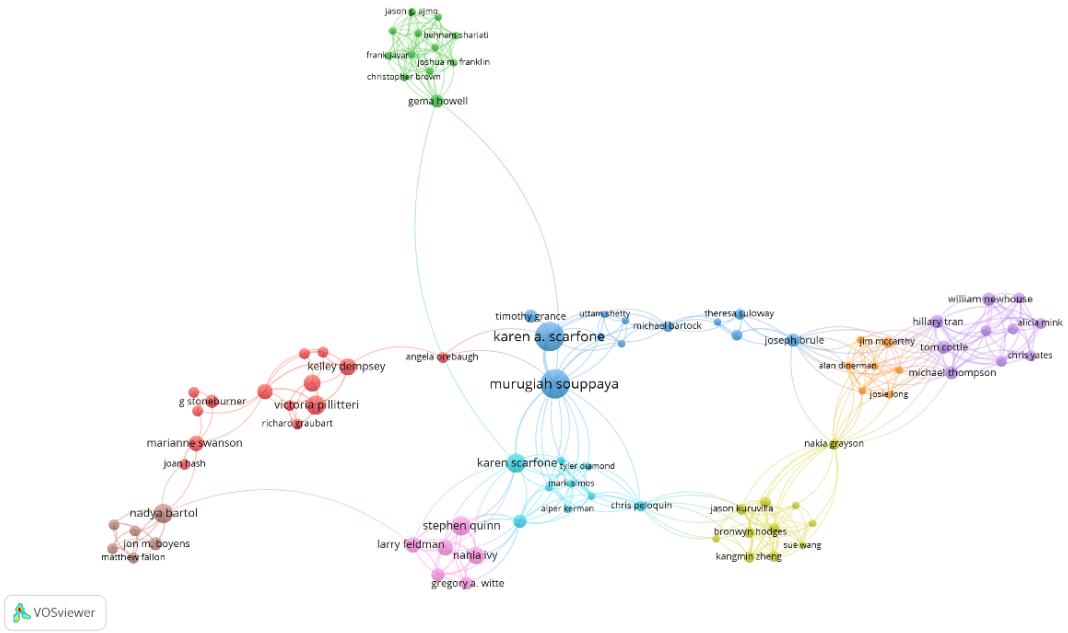

Fig 3. Authors Cluster Map
Source: Authors' elaboration based on VOSviewer outputs and The Lens metadata (2025).

3.1. Results addressing RQ1: Key cybersecurity and network security issues in remote work

The keyword co-occurrence visualization was a main instrument in reorganizing the article's analytical sections. The VOSviewer map highlighted the most frequently used keywords in the articles under review, such as "cybersecurity", "remote work", "risk management", "information security", "Zero Trust", and "endpoint security", which appeared to be the most consolidated clusters (Figure 4). These clusters were a guide for the thematic organization of sections three to six, as well as confirming the topics' relevance, such as identity management, distributed risks, human factors, and control frameworks in remote work

security. Besides, some low-frequency words like “VPN vulnerabilities”, “phishing”, and “home networks” were helpful in pointing out the necessity of technical misconfigurations and user behavior prediction.

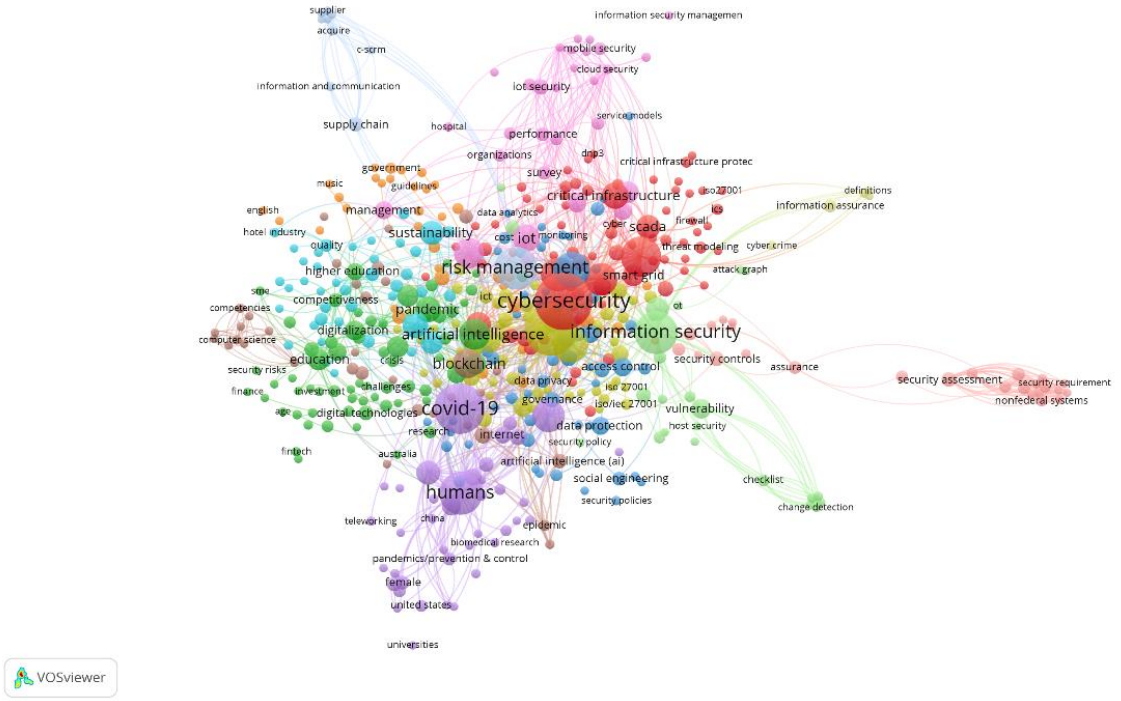

Fig 4. Keywords Cluster Map
Source: Authors’ elaboration based on VOSviewer outputs and The Lens metadata (2025).

Since 2020, organizations have become increasingly dependent on remote, hybrid, and distributed work models, a change that has significantly impacted how organizations operate and the security of their infrastructures. Remote work refers to employees performing professional activities from home, hybrid work is a mix of home-based and office-based activities, and distributed work involves teams operating with varying degrees of autonomy (Bhagat, 2023). These models have reshaped enterprise IT environments and introduce new structural and security challenges.

3.2. Results addressing RQ2: Increased attack surface and vulnerabilities in remote working environments

Human error remains a dominant factor, with studies showing that remote workers comply with security guidelines only partially (Nyarko & Fong, 2023). Home Wi-Fi networks are frequently misconfigured, often relying on old routers or default administrative passwords,

which increases the risk of cyberattacks (Klint, 2023). Mixing personal and corporate digital activities leads to additional concerns about data exposure and unauthorized access.

3.3. Results addressing RQ3: Adequacy of cybersecurity standards and frameworks for remote auditing

Distributed networks introduce vulnerabilities that differ from those typically found in traditional on-premises systems. Misconfigurations of VPN gateways, home routers, and firewalls are usually a result of the fast deployment and lack of oversight in home environments (Treacy et al., 2023). The practice of BYOD further complicates security, as personal devices generally lack enterprise-level protections such as EDR/XDR, centralized patch management, or secure configurations (Bhagat, 2023). Sharing of devices or leaving them unattended in-home settings increases exposure and is a source of physical security risks as well (Klint, 2023).

Lack of proper network segmentation makes it possible for corporate devices to be on the same networks with personal electronics and IoT appliances, which facilitates lateral movement (Silva Atencio, 2025). Security teams are having difficulties with limited visibility, as remote endpoints are often in areas with unstable connectivity or lack standard monitoring tools. Various industry reports reveal that there are persistent difficulties in log forwarding, forensic data collection, and monitoring coverage in non-controlled networks (Deloitte NASCIO Cybersecurity Study, 2024). These gaps in visibility are a direct indication of auditing processes, as auditors frequently report partial evidence and restricted access to remote configurations (Celestin, 2019).

VOSviewer's keyword co-occurrence analysis identified cybersecurity frameworks and auditing procedures as recurrent and significant themes in the bibliometric dataset. Based on these findings, sections three through six were organized according to the thematic structure derived from keyword mapping. While section three explicitly addresses the results related to the research questions, sections four and five provide a more detailed analysis of auditing requirements and the applicability of existing cybersecurity standards, such as ISO/IEC 27001 and the NIST Cybersecurity Framework, in remote and distributed work environments. In this context, although current cybersecurity frameworks remain applicable, their effective implementation increasingly requires the integration of governance-oriented

controls with advanced technical solutions, including Zero Trust architectures and continuous endpoint monitoring tools.

4. Auditing in Remote Work Environments

4.1 Traditional security auditing

Traditionally, security auditing relied on centralized systems, standardized device baselines, and direct access to corporate infrastructures. On-site audits followed structured stages such as defining the scope, reviewing documents, conducting interviews, gathering evidence, and testing configurations (Ilori et al., 2022). The methods used were built on the idea that resources were kept inside a secure perimeter and managed through corporate tools.

Under these models, system logs, configuration files, and event reports were collected from authenticated connections to internal networks. Compliance verification relied heavily on the direct observation of security measures, patch management activities, and intrusion protection. As the infrastructures were centrally managed, auditors were able to work in safe and less challenging contexts.

4.2 Challenges of auditing remote environments

Traditional audit models, which were not designed for decentralized contexts, have been under significant pressure due to remote work. One of the main problems is that it is difficult to get reliable evidence from devices that are operating outside of corporate networks. Remote endpoints may not be permanently connected and enterprise-level monitored, so there may be gaps, delays, and inconsistencies in the data obtained from the audit (Celestin, 2019).

Besides, the lack of visibility is also a major challenge. Device logs are often incomplete and not forwarded to SIEM systems due to bandwidth limitations or misconfigurations. Home networks are generally non-segmented, and even when they are securely configured, auditors find it hard to determine the efficiency of security measures because they must rely on these networks (Klint, 2023). Without reliable telemetry, audit teams hardly find a way to carry out patching, encryption, authentication, and compliance mechanisms.

The problem of user-managed devices has been around for a very long time, and now it has become worse since remote workers might be sharing hardware, turning off security features, postponing updates, or even installing unauthorized software, thus breaking the assumption of controlled configurations and enforced baselines. The authentication and access management processes have been made difficult by the adoption of cloud platforms and federated identity systems. This transition requires MFA, secure channels, and continuous verification. Audits done remotely are still hard when people use outdated VPNs or unsegmented access paths (Allah Rakha, 2023). Patch management is made more difficult by poor or intermittent connectivity leading to devices being left unpatched or outdated (Treacy et al., 2023).

Forensic validation is also confronted with similar issues: severely limited physical or administrative access makes it particularly difficult to confirm logs, check system integrity, and reconstruct incidents. These constraints point to the necessity of modifying audit methods so that they are in line with the distributed infrastructures in contemporary remote work contexts.

4.3 Emerging audit requirements

As companies move to remote and hybrid work styles, the auditing procedures should be changed accordingly. One of the main necessities is utilizing the tools for remote auditing, which should be able to collect the evidence in a standardized, automated manner and be tamper resistant. These are secure remote-access platforms, cloud-integrated dashboards, and endpoint-monitoring systems (Celestin, 2019).

Automation is the main actor in the play. Contemporary auditing is turned more and more towards automated methods of evidence gathering, configuration scanning, and continuous monitoring that can be accomplished using technologies such as EDR/XDR, SIEM, and cloud-native telemetry. The research is quite clear in its findings to support this statement that the automation elevates the exactness of detection and lessens the audit fatigue by on-the-spot identification of anomalies (Ilori et al., 2022).

Accountability for remote work must also be established through proper governance policies. The organizations have implemented policies that regulate the configurations of secure home networks, requirements for endpoint protection, communication by means of

encrypted channels, and procedures for identity verification (Rakha, 2023). The policies must be subject to audit and should be convertible as far as the changes in threats are concerned.

Moreover, it is also very important for the company to be closely linked with the endpoint-security technologies. EDR/XDR technologies offer the transparency necessary to confirm that the latest changes have been made, that the system is free of malware, and that access control requirements have been met, which in turn contributes to the reduction of the auditing evidence shortages (Silva Atencio, 2025).

Considerations related to the human factor should not be forgotten as well. Behavioral risks are intensified by remote work; therefore, auditors need to assess training effectiveness, phishing-resistance programs, and compliance with remote work guidelines (Bhagat, 2023).

In addition, cybersecurity frameworks such as ISO/IEC 27001, NIST CSF, and CIS Controls are putting an ever-greater emphasis on risk-based and continuous monitoring approaches that are very compatible with remote environments (Folorunso et al., 2024). They provide support for adaptable audit models that combine governance reviews, technical controls, and real-time telemetry.

5. Security Frameworks and Their Applicability to Remote Work

The literature obtained from bibliometric research illustrates how cybersecurity frameworks-ISO/IEC 27001, NIST Cybersecurity Framework (CSF), and CIS Controls-have been instrumental in defining security and auditing practices for distributed environments. The central theme, however, revolves around the fact that these frameworks were initially conceptualized for more centralized infrastructures. Still, their fundamental ideas are often mentioned in the documents analyzed, suggesting that they are still viable models for security governance in remote work (Bhagat, 2023).

To sum up, the research works demonstrate that although the present frameworks are still applicable, their efficient implementation in remote work situations requires additional measures and appropriate contextual adaptation. Integrated interpretation, as supported by bibliometric evidence, is a strong argument for the necessity of combining governance, risk-based, and technical layers in remote work security rather than depending on a single framework.

6. Discussion

The keyword co-occurrence analysis (Figure 4) reveals that the most central concepts, remote work, cybersecurity, network security, risk management, and audit, are the main core around which other themes revolve. This suggests that literature is increasingly combining the aspects of technical risks, access control, and auditing practices for the formation of a single conceptual framework. The emergence of several clusters also implies that there is a certain degree of fragmentation in the field, with some research focusing on technical vulnerabilities and others emphasizing organizational and human-factor aspects. In addition, the bibliometric evidence suggests relevant geographical and institutional differences in how remote work cybersecurity is approached. Publications originating from North America and Northern Europe tend to emphasize technical controls, automation, and Zero Trust architectures, while studies from other regions place greater focus on regulatory compliance, privacy, and legal constraints. This divergence reflects distinct organizational cultures, regulatory environments, and maturity levels in cybersecurity governance. From an auditing perspective, these differences highlight the need for flexible audit frameworks that can accommodate both highly automated technical environments and compliance-driven contexts without compromising assurance quality.

7. Conclusions

The bibliometric review offered through this article provides a comprehensive understanding of how remote work has changed network security risks, as well as the auditing practices that are necessary to tackle them. The Lens dataset and VOSviewer visualizations (Figures 1 and 2) provided the results that depicted a research landscape influenced by operational, regulatory, and technological pressures.

References

Bhagat, N. (2023). Cybersecurity in a Remote Work Era: Strategies for Securing Distributed Workforces. *Review Article Nikhil Bhagat, Sr. Technical Account Manager Network Specialist Independent Scholar*, *2*(2), 1–5. https://doi.org/10.47363/JMCA/2023(2)E137

Celestin, M., & Vanitha, N. (2019). The rise of remote auditing: Challenges, opportunities, and best practices. *International Journal of Computational Research and Development*, *4*(2), 13-20.

Deloitte NASCIO Cybersecurity Study. (2024). https://www.deloitte.com/us/en/ Insights/industry/government-public-sector-services/2024-deloitte-nascio-cybersecurity-study.html

Folorunso, A., Mohammed, V., Wada, I., & Samuel, B. (2024). The impact of ISO security standards on enhancing cybersecurity posture in organizations. *World Journal of Advanced Research and Reviews*, *24*(1), 2582–2595. https://doi.org/10.30574/wjarr.2024.24.1.3169

Ilori, O., Lawal, C. I., Friday, S. C., Isibor, N. J., & Eke, E. C. C.-. (2022). Cybersecurity Auditing in the Digital Age: A Review of Methodologies and Regulatory Implications. *Journal of Frontiers in Multidisciplinary Research*, *3*(1), 174–187. https://doi.org/10.54660/.ijfmr.2022.3.1.174-187

Klint, R. (2023). Cybersecurity in Home-Office Environments. *DiVA Portal*, 53. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1779054&dswid=-4035

Nyarko, D. A., & Fong, R. C.-W. (2023). Cyber security compliance among remote workers. Em *Advanced Sciences and Technologies for Security Applications* (pp. 343–369). Springer International Publishing.

Rakha, N. A. (2023). Ensuring cyber-security in remote workforce: legal implications and international best practices. *International Journal of Law and Policy*, *1*(3), 1-19.

Silva Atencio, G. (2025). Effective Cybersecurity Strategies for Mitigating Remote Work and IoT Risks in Enterprises. *FinTech and Sustainable Innovation*. https://doi.org/10.47852/bonviewfsi52025962

Treacy, S., Sabu, A., Bond, T., O'Sullivan, J., Sullivan, J., & Sylvester, P. (2023, February). Organizational cybersecurity post the pandemic: an exploration of remote working risks and mitigation strategies. In *International Conference on Cyber Warfare and Security* (Vol. 18, No. 1, pp. 394-401). Academic Conferences International Limited.

https://doi.org/10.58086/r7pn-1f89

SECURITY AND PRIVACY IN EXPLAINABLE AI: A BIBLIOMETRIC ANALYSIS OF EMERGING LEAKAGE RISKS

Mafalda Matos[1]✉, **Mário Dias Lousã**[1,2], **José Carlos Morais**[1,3]

[1] Instituto Superior Politécnico Gaya (ISPGAYA), Portugal.
[2] Insight - Piaget Research Center for Ecological Human Developement, Portugal.
[3] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal.
✉ Corresponding authors**:** ispg2024103142@ispgaya.pt

Abstract

Explainable Artificial Intelligence (XAI) has gained increasing attention as a means of improving the transparency and trustworthiness of machine learning algorithms, particularly in domains where security and privacy concerns are relevant. This study presents a bibliometric analysis of research at the intersection of explainable artificial intelligence, security, and privacy. The aim was to characterize publication trends, thematic structures, and keyword relationships within the field. Scholarly records were retrieved from the Lens database using a structured search strategy based on the PRISMA protocol and analyzed using bibliometric tools, including Bibliometrix and VOSviewer. The total number of studies analyzed was 8,099, and the analyzed time frame was 2010–2025. The analysis examined general publication information, annual scientific production, leading publication venues, and keyword co-occurrence networks. Results indicate a rapid growth in XAI-related publications in recent years and reveal several major thematic clusters, including deep learning–driven medical imaging applications, foundational machine learning and data science concepts, explainability methods in security and distributed learning contexts, and governance-oriented themes related to ethics, privacy, and trust. Overall, the findings highlight the application-driven and interdisciplinary nature of explainable AI research, while showing that security and privacy topics, although present, remain relatively peripheral within the broader XAI literature.

Keywords: Explainable Artificial Intelligence; Bibliometric Analysis; Cybersecurity; Information Leakage; Machine Learning; Data Science; Privacy-Preserving ML.

1. Introduction

The increasing adoption of machine learning in security-sensitive and high-stakes domains has intensified interest in explainable artificial intelligence (XAI) as a means of improving transparency, trust, and accountability in automated decision-making systems (Doshi-Velez & Kim, 2017). In areas such as cybersecurity and related application domains, explainability is frequently promoted to support analyst decision-making, facilitate regulatory compliance, and enhance the interpretability of complex detection models (Capuano et al., 2022). As a result, research at the intersection of XAI, security, and privacy has expanded rapidly over the past decade.

At the same time, previous studies have revealed potential conflicts between explainability and security. By revealing information about model behavior or internal decision logic, explanation mechanisms may inadvertently expose additional information that could be exploited by adversaries, raising concerns related to privacy and information leakage (Shokri et al., 2017). While these issues are increasingly acknowledged, existing research remains dispersed across multiple disciplines and application domains, complicating efforts to obtain a consolidated view of how security and privacy concerns are positioned within the broader XAI literature.

In this context, bibliometric analysis offers a systematic and quantitative approach to mapping publication patterns, thematic structures, and research trends across large and interdisciplinary bodies of literature (Donthu et al., 2021).

Accordingly, this study conducts a bibliometric analysis of scholarly research at the intersection of explainable artificial intelligence, security, and privacy using data retrieved from the Lens database. The analysis examines publication growth, source distribution, and keyword co-occurrence patterns to characterize the conceptual structure of the field, with results presented in section four and discussed in section five. The objective of this work is not to evaluate specific algorithms or security mechanisms but to provide a high-level overview of research trends and thematic emphasis that can support future systematic reviews and empirical investigations.

To guide the bibliometric analysis, this study is structured around the following research questions:

RQ1: How has scientific production related to explainable artificial intelligence, security, and privacy evolved over time?

RQ2: Which journals and publication venues have contributed most prominently to research at the intersection of explainable AI, security, and privacy?

RQ3: How are security- and privacy-related themes positioned within the broader explainable AI research landscape?

2. Conceptualization

2.1. Lightweight Machine Learning in Security Applications

Lightweight machine learning models are widely adopted in security-sensitive and resource-constrained environments due to their low computational overhead and relatively transparent decision structures. In application domains such as intrusion detection, malware analysis, and Internet of Things (IoT) security, models including logistic regression, decision trees, random forests, and support vector machines are commonly employed to balance predictive performance with deployment feasibility on edge and embedded systems (Buczak & Guven, 2016; Khraisat et al., 2019). Their continued prevalence is driven by practical constraints related to processing power, memory, energy consumption, and the need for timely operational decisions (Shi et al., 2016). From an explainability perspective, the structural simplicity of these models facilitates post-hoc interpretation and transparency, making them frequent baselines in explainable artificial intelligence research within applied security and privacy contexts (Molnar, 2022).

2.2. Explainable Artificial Intelligence

Explainable Artificial Intelligence (XAI) encompasses methods and techniques aimed at making machine learning model outputs more transparent and interpretable to human users, particularly as such systems are increasingly deployed in high-stakes domains (Floridi et al., 2018; Gunning et al., 2019). Existing approaches are commonly distinguished between model-intrinsic methods and post-hoc explanation techniques applied to black-box models, with explanations further characterized as global or local in scope (Guidotti et al., 2018). Widely adopted post-hoc methods include feature attribution and surrogate modeling approaches that approximate model behavior to support human understanding (Ribeiro et al., 2016). In applied contexts such as healthcare, finance, and cybersecurity, explainability is

typically framed as a mechanism for supporting human oversight, regulatory compliance, and operational trust rather than as a replacement for automated decision-making (Samek et al., 2021). From a bibliometric perspective, the diversity of definitions, methods, and application domains has contributed to a fragmented and interdisciplinary research landscape.

2.3. Explainable Artificial Intelligence in Security and Privacy Contexts

Explainable artificial intelligence has attracted growing attention in security and privacy-sensitive domains, where understanding and justifying automated decisions is often as critical as predictive accuracy (Adadi & Berrada, 2018). In cybersecurity applications, XAI techniques are frequently employed to support intrusion detection, malware classification, anomaly detection, and fraud analysis by providing interpretable insights into model behavior for analysts and decision-makers (Amershi et al., 2014; Tramèr et al., 2016). While explainability is primarily motivated by practical considerations such as alert validation, trust-building, and operational transparency, its integration into machine learning pipelines also raises potential security and privacy concerns. Prior work has shown that explanation outputs may inadvertently expose information about model internals or training data, particularly in adversarial or untrusted environments (Tramèr et al., 2016). From a bibliometric standpoint, this dual role of XAI, as both a transparency-enhancing mechanism and a potential source of additional risk, contributes to the dispersion of security and privacy-related themes across multiple research clusters.

2.4. Conceptual Positioning and Scope of the Bibliometric Analysis

The combined proliferation of lightweight machine learning models, the methodological diversity of explainable artificial intelligence, and the increasing use of XAI in security and privacy-sensitive contexts have resulted in a rapidly expanding and heterogeneous body of research spanning multiple disciplines and application domains (Adadi & Berrada, 2018; Samek et al., 2021). While prior studies provide valuable conceptual frameworks and application-specific insights, the fragmentation of literature complicates efforts to identify dominant themes and broader research trajectories. In this context, bibliometric analysis offers a

complementary, high-level perspective by enabling systematic mapping of publication patterns, thematic structures, and the relative positioning of security and privacy concerns within the broader explainable AI landscape (Donthu et al., 2021). Rather than evaluating individual methods, this approach supports the identification of research trends and emerging gaps across interdisciplinary bodies of work.

3. Methodology

A bibliometric analysis was conducted to map the structure and evolution of research at the intersection of explainable artificial intelligence, security, and privacy. The analysis followed standard bibliometric procedures, including corpus construction, descriptive indicator analysis, and keyword co-occurrence mapping, and was conducted in December of 2025.

More specifically, the parameters used were:

- Years: 2010–2025.
- Language: English.
- Document types: Articles, Reviews, Conference papers.
- Database: The Lens.
- Documents analyzed: All that fit the previous criteria.

As Lens does not provide a dedicated language metadata field, English-language publications were identified through abstract-based filtering, followed by manual validation of a random sample. This approach is consistent with prior bibliometric studies using Lens.

These were the exact search parameters used:

Scholarly Works (8,198) = (("explainable AI" OR (XAI OR ("interpretable machine learning" OR "model interpretability"))) AND (“security” OR (“cybersecurity” OR (privacy OR "information leakage"))))

Filters: Year Published = (2010 -)

Publication Type = (journal article , review , journal issue , conference proceedings , journal , conference proceedings article) Field of Study = (Computer science , Artificial intelligence , Machine learning , Data science , Computer security)

A PRISMA-inspired workflow (Figure 1) was employed to transparently report the identification and preprocessing of records for bibliometric analysis. The workflow reflects

automated filtering and metadata-based preprocessing steps. No further steps were taken for manual verification.

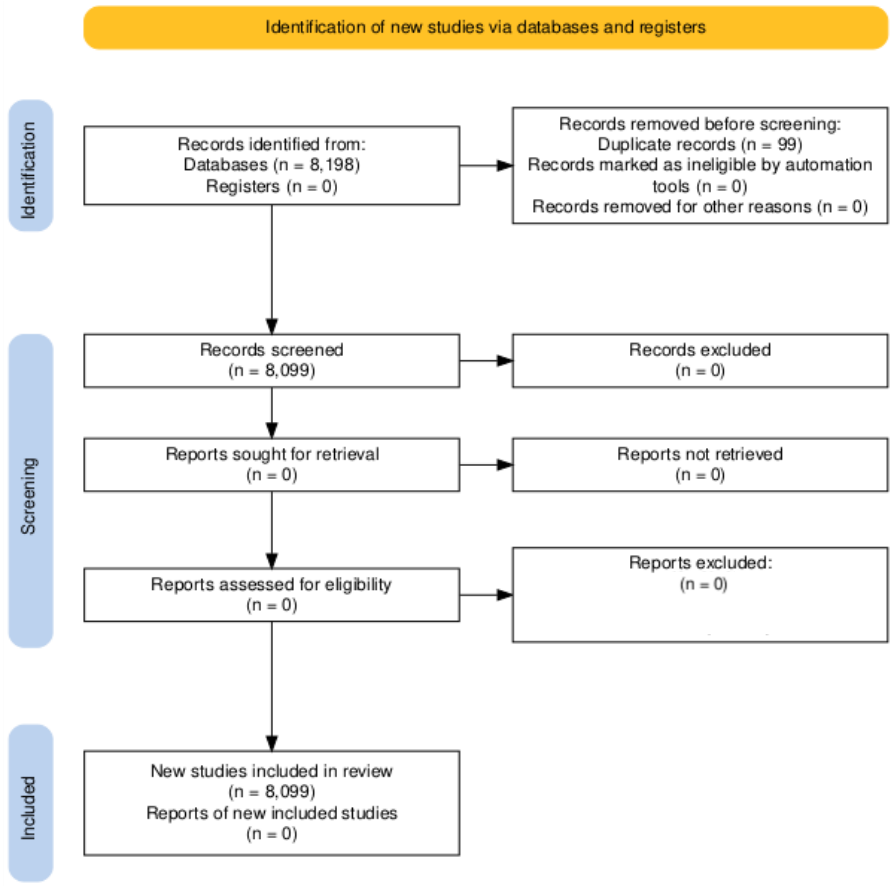

Fig 1. PRISMA-inspired workflow schema
Source: Own generation using https://estech.shinyapps.io/prisma_flowdiagram/

4. Results and Discussion

4.1. Overall Dataset Characteristics

The final bibliometric corpus comprises 8,099 publications indexed in the Lens database, spanning the period 2010–2025. These publications are distributed across 2,669 distinct sources, reflecting the highly multidisciplinary nature of research at the intersection of explainable artificial intelligence, security, and privacy. The dataset exhibits a very high annual growth rate (59.7%). The average document age is 1.57 years, highlighting the strong recency bias of literature.

4.2. Temporal Evolution of Publications

Analysis of annual scientific production (Figure 2) reveals a slow emergence phase between 2010 and 2016, followed by a sharp acceleration starting in 2018. Publication counts increase significantly after 2019, with particularly strong growth observed from 2021 onward. The most pronounced rise occurs in 2024 and 2025, which together account for a substantial proportion of the total corpus.

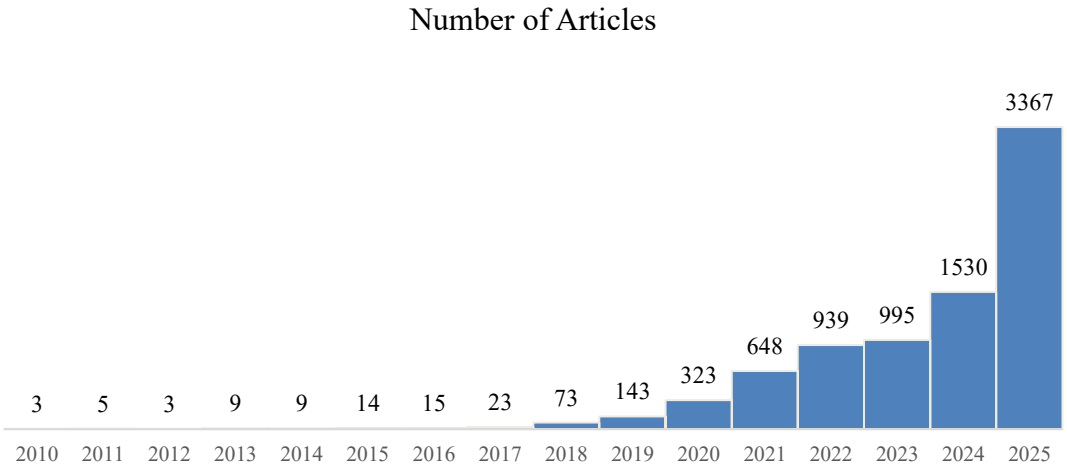

Fig 2. Evolution of the number of publications on explainable artificial intelligence, security, and privacy (2010–2025)
Source: Authors' elaboration based on Bibliometrix and python.

This surge coincides with the widespread adoption of deep learning systems and the increasing demand for explainability in high-stakes domains, suggesting that concerns related to transparency, accountability, and privacy have become central drivers of recent research activity.

4.3. Source Distribution and Publication Venues

The literature is highly dispersed across venues, with no single dominant outlet, further reinforcing the multidisciplinary character of the field (Table 1 and Figure 3). The sources with the highest number of publications address applied artificial intelligence and data-driven decision-making.

In terms of publications, the strong prevalence of biomedical, healthcare, and sensing-oriented journals among the top-ranked sources (Table 1 and Figure 3). This suggests that security and privacy issues related to XAI are most frequently investigated within data-sensitive application domains, such as medical diagnosis, health monitoring, and clinical decision support, rather than in traditional cybersecurity venues.

Table 1. Top 20 publication venues by number of documents in the analyzed corpus

Source	Number of Publications
Scientific Reports	606
Sensors	359
PLOS ONE	151
Diagnostics	146
IEEE Access	128
Frontiers in Artificial Intelligence	120
Applied Sciences	100
Journal of Medical Internet Research	95
PeerJ Computer Science	93
Bioengineering (Basel, Switzerland)	89

Source: Authors' elaboration based on data from The Lens database (2025)

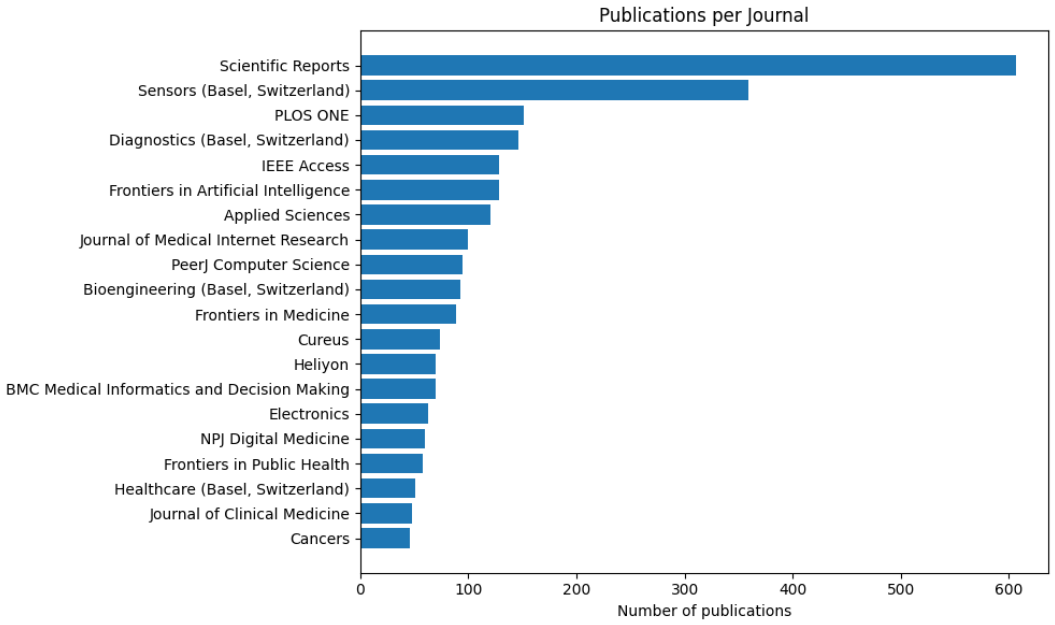

Fig 3. Top 20 publication venues by number of documents in the analyzed corpus
Source: Bibliometrix and python

4.4. Authorship and Collaboration Patterns

The dataset includes contributions from 35,922 unique authors, with an average of 5.08 co-authors per document, indicating a predominantly collaborative research culture. Single-authored publications account for only 887 documents, reflecting the interdisciplinary nature of XAI-related research.

The most productive authors are distributed across diverse institutional and disciplinary backgrounds, reflecting the absence of a small, centralized research community. Instead, contributions emerge from a broad range of applied AI, machine learning, and domain-specific research groups.

4.5. Keyword Coverage and Metadata Characteristics

Figure 4 presents the keyword co-occurrence network, with a minimum occurrence limit of 30 to balance thematic scope and readability, resulting in 78 keywords.

The analysis reveals several dominant application-driven clusters, particularly in medical imaging, healthcare analytics, and data-intensive biomedical domains. Explainable AI methods form a distinct cluster connected to anomaly detection, IoT, and intrusion detection, while security and privacy-related terms appear comparatively smaller and more peripheral. This suggests that, although security and privacy concerns are present in XAI research, they are not yet central organizing themes within the broader literature.

Given the highly fragmented and application-driven nature of the field, the analysis focuses on thematic and venue-level structures rather than individual author productivity.

A thematic analysis (not shown) further supports the findings of the keyword co-occurrence network. Dominant themes are centered on applied machine learning and explainability in healthcare and sensing contexts, while security and privacy-related themes appear as less central and less developed. This thematic configuration reinforces the observation that security considerations in XAI research remain secondary to application-driven concerns.

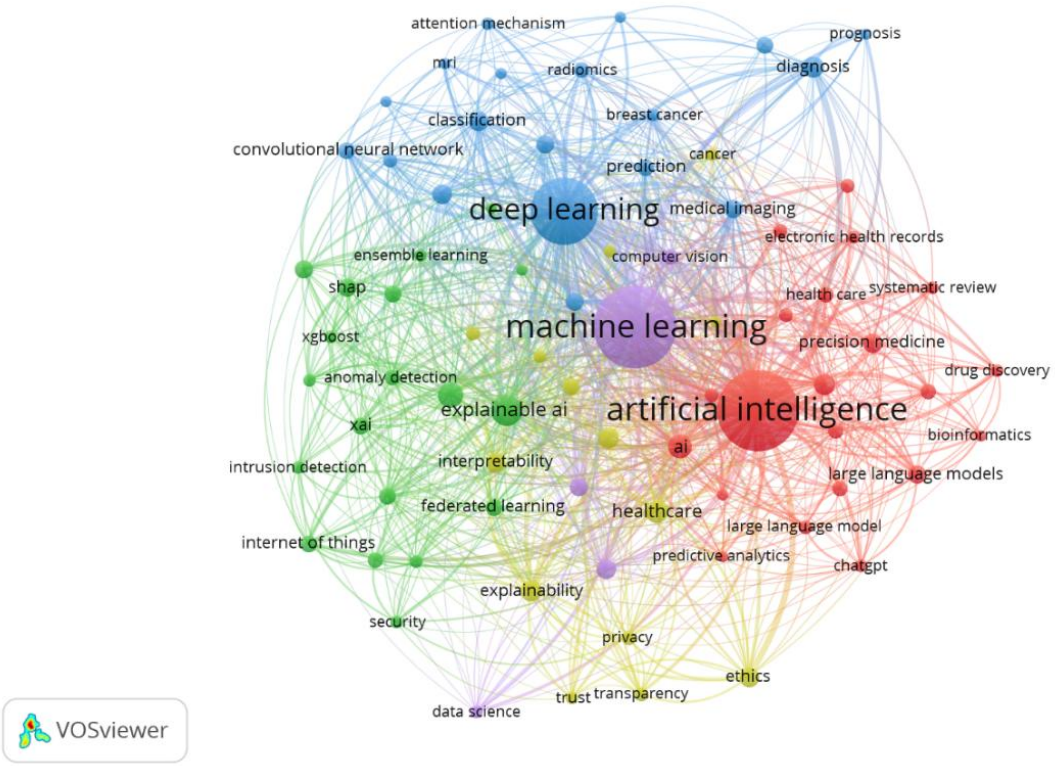

Fig 4. Keyword Co-occurrence Network Figure
Source: Own generation in VOSViewer

Purple Cluster: Core Machine Learning and Data Science Foundations

The purple cluster represents the methodological core of the explainable AI literature and is characterized by foundational terms such as machine learning, computer vision, and data science. The centrality and strong interconnectivity of these keywords indicate their role as shared technical foundations across diverse application areas rather than as a distinct domain of inquiry. The prominence of computer vision reflects the importance of visual data processing tasks within the corpus, particularly in contexts where model performance and interpretability intersect. Overall, this cluster functions as a structural backbone linking application-driven research with more specialized methodological and thematic clusters, including explainability techniques and security-related topics.

Red Cluster: General Artificial Intelligence and Data-Driven Application Domains

The red cluster captures a broad, application-oriented body of research centered on general artificial intelligence concepts and data-driven methodologies. Core keywords such as artificial intelligence and machine learning co-occur with terms associated with contemporary

AI paradigms, including large language models, predictive analytics, and generative systems. The strong presence of healthcare-related applications—such as precision medicine, bioinformatics, drug discovery, and electronic health records—highlights the dominant role of biomedical domains within the broader AI literature captured in the dataset. The inclusion of systematic review further suggests a maturing research stream characterized by rapid growth and increasing synthesis activity.

Blue Cluster: Deep Learning–Driven Medical Imaging and Diagnostic Applications

The blue cluster is dominated by deep learning–based medical imaging research, with prominent keywords including deep learning, convolutional neural networks, classification, and attention mechanisms. The strong association with clinical terms such as breast cancer, radiomics, MRI, diagnosis, and prognosis indicates a concentration on imaging-based decision-support systems in healthcare settings. Within this cluster, advanced neural architectures are commonly applied to high-dimensional visual data to support disease detection and outcome prediction. The presence of attention mechanisms reflects ongoing efforts to enhance both predictive performance and interpretability in clinically oriented deep learning applications.

Green Cluster: Explainable AI Methods in Security and Distributed Learning Contexts

The green cluster centers on explainable artificial intelligence methods applied in security and infrastructure-oriented contexts. Key terms such as explainable AI, XAI, and SHAP co-occur with intrusion detection, anomaly detection, and security, indicating the use of explanation techniques to support transparency and analyst interpretation in detection systems. The presence of ensemble methods and models such as XGBoost reflects a balance between predictive performance and post-hoc interpretability. In addition, keywords related to IoT and federated learning highlight the relevance of distributed and resource-constrained environments, where explainability is used to support trust and insight without centralizing sensitive data.

Yellow Cluster: Interpretability, Trust, and Ethical Considerations in Applied AI

The yellow cluster encompasses interpretability and governance-oriented themes, including explainability, privacy, ethics, trust, and transparency. These keywords reflect a body of literature concerned with the societal, regulatory, and operational implications of deploying

machine learning systems in sensitive contexts. The frequent association with healthcare-related terms underscores the prominence of ethical and trust-based discussions in clinical and biomedical applications, where regulatory compliance and patient safety are central concerns. Rather than emphasizing specific technical methods, this cluster highlights normative and governance perspectives that cut across application domains and connect explainability research with broader discussions of responsible and trustworthy AI.

4.6. Security and Information Leakage Implications of Explainable AI

The results indicate that research on explainable artificial intelligence at the intersection with security and privacy is predominantly application-driven and dispersed across multiple domains and publication venues. Although security- and privacy-related concerns are present throughout the analyzed corpus, they are primarily embedded within applied contexts—particularly in the healthcare and Internet of Things domains—rather than being articulated from explicitly adversarial or threat-model–centric perspectives. The bibliometric analysis reinforces this observation by showing that keywords associated with security and privacy do not play a structuring role in the thematic organization of the explainable artificial intelligence literature. Notably, terms related to concrete information leakage mechanisms and attack models occupy peripheral positions within the keyword co-occurrence network, suggesting that the potential of explainability mechanisms to introduce new attack surfaces has not yet been systematically addressed at the landscape level. This peripheral positioning highlights an emerging research gap at the intersection of explainability, security, and privacy, in which information leakage risks induced by explanation methods remain underexplored.

5. Conclusions, Limitations and Future Work

This study presented a bibliometric analysis of research intercepting explainable artificial intelligence, security, and privacy, aiming to map publication trends, thematic structures, and keyword relationships within this rapidly expanding field. By analyzing a large corpus of publications indexed in the Lens database, the study provides a structured overview of

how explainability-related research has evolved over time and how security and privacy-oriented topics are situated within the broader XAI literature.

The results indicate a sharp growth in scientific output in recent years, reflecting the increasing relevance of explainable AI across multiple application domains. Keyword co-occurrence and thematic analyses reveal that explainable AI research is strongly intertwined with general machine learning and AI concepts, with deep learning playing a central role. At the same time, security- and privacy-related topics appear as more peripheral but increasingly connected themes, often embedded within applied contexts rather than forming standalone research clusters.

Overall, this bibliometric mapping highlights the interdisciplinary and application-driven nature of explainable AI research while providing a high-level quantitative overview that can support more focused systematic reviews and empirical investigations.

This study has several limitations that should be considered when interpreting the results. First, the analysis relies on metadata retrieved from a single bibliographic database. Although the Lens platform offers broad coverage across disciplines, some relevant publications indexed exclusively in other databases may not be included.

Second, the bibliometric approach is inherently dependent on the quality and consistency of bibliographic metadata, particularly author keywords. Variations in terminology, keyword granularity, and indexing practices may influence the structure of the identified clusters and co-occurrence networks.

Third, bibliometric methods capture patterns of publication activity and thematic relationships but do not assess the methodological quality, empirical validity, or practical effectiveness of individual studies. As a result, the findings should be interpreted as indicative of research trends rather than as evaluative judgments about specific approaches or techniques.

Finally, the dominance of certain application domains, such as healthcare, reflects the distribution of the existing literature and does not imply that explainable AI research is inherently limited to these contexts. The scope of this work was kept general on purpose to capture the full landscape of this field, which necessarily means it emphasizes coverage over depth, and future work should explore domain-specific nuances in greater detail.

The findings also point to differentiated implications for public and private stakeholders. From a regulatory perspective, the prominence of ethics and governance-related themes suggests that transparency and explainability requirements should be coupled with explicit consideration of security and information leakage risks, particularly in adversarial or high-stakes

contexts. From a product development perspective, explainability mechanisms should be treated as potential risk vectors rather than purely trust-enhancing features, requiring integration into security and privacy-by-design practices. This distinction highlights the need to align ethical and governance-oriented discussions of explainable AI with concrete deployment and threat considerations.

Future research could extend this bibliometric analysis in several directions. First, integrating multiple bibliographic databases may provide a more comprehensive representation of the explainable AI research landscape and reduce potential database-specific biases.

Second, the descriptive insights presented in this study could serve as a foundation for systematic literature reviews focusing on specific subtopics, such as explainability techniques in adversarial settings, privacy-preserving explanations, and information leakage risks introduced by explainability mechanisms, particularly in security- and privacy-sensitive deployment contexts. Such reviews would enable deeper qualitative analyses that go beyond the scope of bibliometric mapping.

Third, longitudinal analyses examining the evolution of security and privacy-related themes over shorter time intervals could provide finer-grained insights into emerging research directions.

Finally, future work may explore the relationship between explainability research and regulatory or governance frameworks, particularly as AI becomes an increasingly prominent policy concern.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine, 35*(4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

Capuano, N., Fenza, G., Loia, V., & Stanzione, C. (2022). Explainable artificial intelligence in CyberSecurity: A survey. *IEEE Access: Practical Innovations, Open Solutions*, *10*, 93575–93600. https://doi.org/10.1109/access.2022.3204171

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. arXiv preprint arXiv:1702.08608. https://arxiv.org/abs/1702.08608

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), 1–42. https://doi.org/10.1145/3236009

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI-Explainable artificial intelligence. *Science robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis Campbell Systematic Reviews, 18, e1230. https://doi.org/10.1002/cl2.1230

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity, 2*(1), Article 20. https://doi.org/10.1186/s42400-019-0038-7

Molnar, C. (2022). *Interpretable machine learning* (2nd ed.). Leanpub. https://christophm.github.io/interpretable-ml-book/

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Samek, W., Wiegand, T., & Müller, K.-R. (2021). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *IEEE Signal Processing Magazine, 38*(3), 40–48. https://doi.org/10.48550/arXiv.1708.08296

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal, 3*(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *Proceedings of the IEEE Symposium on Security and Privacy* (pp. 3–18). https://doi.org/10.1109/SP.2017.41

Tramèr, F., Zhang, F., Juels, A., Reiter, M. K., & Ristenpart, T. (2016). Stealing machine learning models via prediction APIs. In *Proceedings of the 25th USENIX Security Symposium* (pp. 601–618).

https://doi.org/10.58086/hbgd-e714

MALVERTISING AS A VECTOR OF CYBERCRIME IN DIGITAL PLATFORMS

João Sousa[1]✉, **Mário Dias Lousã**[1,2], **José Carlos Morais**[1,3]

[1] Instituto Superior Politécnico Gaya (ISPGAYA), Portugal.
[2] Insight - Piaget Research Center for Ecological Human Developement, Portugal.
[3] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal..
✉ Corresponding authors**:** ispg2024104121@ispgaya.pt

Abstract

The usage of malvertising, which exploits advertising networks to send malware, ransomware, and phishing techniques while evading traditional security measures, has become a significant avenue for criminality on digital platforms. This study provides a bibliometric analysis of 236 papers from The Lens database, focusing on the evolution of scientific output, author collaboration patterns, and theme frameworks in malvertising research. Using Bibliometrix (R) and VOSviewer, co-authorship networks, keyword co-occurrence maps, theme clusters, and a density metadata table were created to identify research trends and knowledge gaps. Results show that machine learning, behavioral analytics, and ecosystem-aware security solutions are receiving more attention in the field of research, with high-impact publications like IEEE Access, Sensors, and Electronics making substantial contributions to their development. Additionally, the report highlights research prospects and difficulties by identifying upcoming issues such as blockchain, IoT, AR/VR platforms, and zero-day malvertising. Finally, this report summarizes the state of the art in malvertising research, highlights systemic weaknesses in advertising ecosystems, and offers suggestions for future cybersecurity research topics, adaptive defense tactics, and regulation.

Keywords: Malvertising, Online Advertising Ecosystems, Cybercrime, Malware Distribution, Bibliometric Analysis.

1. Introduction

Digital advertising has become a dominant method of distributing publicity over the past two decades, attracting substantial revenue and integrating deeply into online ecosystems (Sadeghpour & Vlajic, 2021). Due to its growth and scale, it has also become a target for

cybercriminals, who exploit ad networks to deliver malicious content and compromise users (Nowroozi et al., 2022). Malvertising has emerged as a key vector for cybercrime campaigns, where attackers embed ads and exploit kits to bypass traditional security controls (Anand et al., 2023). With the development of mobile applications, the progression of technology further extends the threat, as some apps deliver malicious ads or redirect users to harmful content (Liu et al., 2020). On social media platforms, attackers disguise themselves as legitimate advertisers to target users with harmful ads, leveraging micro-targeting to maximize exposure (Arrate et al., 2020). Despite the rapid growth of research on digital advertising threats, current literature presents malvertising as a vector for cybercrimes (Pooranian et al., 2021). Studies highlight vulnerabilities across the entire advertising ecosystem, including publishers, exchanges, and DSP/SSP networks, yet systematic analyses of prevalence, mechanics, and technological enablers remain scarce (Chua et al., 2020). Recent reviews highlight that malvertising is an initial access technique and is able to partake in complex exploit-kit supply chains, but comprehensive mappings of trends, thematic structures, and author networks are still lacking (Caviglione et al., 2021). This gap defines the need for a bibliometric approach to identify research clusters, emerging topics, and patterns of collaboration in malvertising studies.

A bibliometric analysis was chosen for this study, as it favors an approach where the author examines the scientific research done about malvertising. By analyzing publication patterns, author collaborations, and keyword trends, this method allows for a clear mapping of research activity and thematic development in the field. To ensure the transparency and reproducibility of the process for selecting and screening relevant studies through the recurrent application of the PRISMA 2020 framework. With the usage of bibliometric techniques, it's expected that it will be possible to extract results about the current state of research while highlighting general trends and structures without overextending the available data. The main objective of this study is to analyze the scientific literature on malvertising as a vector of cybercrime in digital platforms using bibliometric techniques, with a particular focus on publication trends, thematic structures, and the evolution of research over time. To achieve this goal, the study addresses the following research questions:

RQ1: How has the scientific production on malvertising evolved over time?

RQ2: What author clusters and collaboration patterns can be identified in malvertising research?

RQ3: What are the main keyword clusters and conceptual structures shaping malvertising research?

RQ4: What are the dominant research themes and topics addressed in the literature on malvertising?

RQ5: How is the scientific literature on malvertising distributed in terms of research density and thematic concentration?

This paper is composed of six sections as follows. In section one, there is an introduction to the study, contextualizing malvertising as a critical vector of cybercrime in digital platforms. Section two provides a theoretical conceptualization of malvertising, outlining its main definitions, mechanisms, and relationship with cybercrime and the way cybercriminals act. The bibliometric methodology used is presented in section three, including data sources, search strategy, and analytical techniques. Section four presents the results of the bibliometric analysis, including author and keyword clusters, publication trends, and density maps. Following the results retrieved from section four, we have section five, where the discussion of the main findings and their implications for research and practice is done. Finally, section six concludes the paper by summarizing the results, highlighting limitations, and suggesting directions for future research.

2. Theoretical Conceptualization

2.1. Malvertising

In cybercrime, there are several ways to attack users, such as malvertising, where malicious code is injected into legitimate advertising content, or a connection is introduced that takes the user to a third-party website, enabling attackers to deliver malware without the user's knowledge (Pooranian et al., 2021). It is particularly effective because it operates within environments that users already view as trustworthy. The attackers responsible for these types of crimes often navigate into popular websites, creating automated connections between users and malware distribution sites, a tactic frequently leveraged in malvertising campaigns (Shin et al., 2022). Unlike traditional attack vectors that rely on explicit user interaction, malicious advertisements are delivered through browsing activities, reducing user suspicion and increasing exposure to attacks. This type of attack typically employs

obfuscated scripts and invisible iframes to redirect users to exploit kits or malware-hosting domains, evading conventional detection systems (Shin et al., 2022). Such operations follow a standard pipeline in which an advertisement leads to a redirect chain, which then directs the user to an exploit kit that delivers a tailored payload, allowing attackers to customize malware delivery based on the user distraction or not having knowledge about the danger without requiring user consent (Caviglione et al., 2021). Such characteristics make this type of cybercrime difficult to detect and mitigate, as the contents are often indistinguishable from legitimate advertising until the attack and proceeding exploitation occur.

2.2. Digital Advertising Ecosystem

The modern online advertising ecosystem is composed of multiple intermediaries, including ad networks, ad exchanges, supply-side platforms, demand-side platforms, and data management platforms, which connect publishers, advertisers, and users (Sadeghpour & Vlajic, 2021). With the fact that digital advertising is highly automated by nature, there are further complications in the case of security enforcement, as decisions related to ad placement and delivery are often executed in milliseconds with minimal human oversight. While these intermediaries enable efficient targeting and monetization, they also expand the attack surface for fraud and malvertising (Sadeghpour & Vlajic, 2021). Also, there are limits to the ability to conduct thorough security validation at each stage of the advertising process due to the speed and efficiency of security checks and the revenue generation priority. This increases the dangers for ad viewers. With the rise in the topic of fraud, there was a necessity to create categories such as placement fraud, traffic fraud, and action fraud, with malvertising being categorized as a key form of placement fraud (Sadeghpour & Vlajic, 2021). Lack of transparency in ad brokerage processes and the speed of real-time bidding systems further complicate creative vetting, allowing malicious ads to bypass security checks (Caviglione et al., 2021). As a result of the current framework, security responsibilities are frequently distributed across multiple employees of the ad platforms, creating gaps that can be exploited by malicious actors.

2.3. Cybercrime Campaigns via Malvertising

With the rise of cybercrime operations, the malvertising in the multiple vectors the cybercriminals use to attack and steal from the multiple digital platform users is done with such strategic thinking. Nowadays most of the used attack vectors are shifting toward low-interaction, high-reach attack models. Malvertising is increasingly being used in broader cybercrime campaigns, including ransomware, phishing, and crypto mining operations (Anand et al., 2023). To achieve a large-scale malware distribution with low maintenance, the cybercriminals use the fact that the advertising channels don't have high levels of security and ad checks in a way to create direct engagement with victims. Attackers exploit advertising infrastructure to deliver malware payloads while bypassing traditional security controls, using malicious ads as covert delivery channels (Nowroozi et al., 2022). Also, the development of mobile applications further extends this threat, delivering harmful ads or redirecting users to malicious content (Liu et al., 2020). Social media platforms amplify the risk, as attackers masquerade as legitimate advertisers or deploy automated bots to distribute fake advertisements and manipulate engagement metrics (Mbona & Eloff, 2021). This framework that the platforms use not only increases campaign efficiency but also complicates attribution and disruption efforts, as malicious activity is embedded within legitimate commercial infrastructures. Malvertising campaigns also exploit weaknesses in digital platforms using plugins, leveraging user trust and executing plugins to take over the privileges of those who fall and are infected to trigger ad injections or redirects across multiple pages (Kasturi et al., 2022).

2.4. Security Implications and Detection Challenges

The usage of malicious advertisement URLs has become challenging to detect because they often mimic legitimate sites and evolve rapidly, rendering traditional detection methods ineffective (Nowroozi et al., 2022). One of the primary difficulties in addressing malvertising lies in the transient nature of malicious advertising content, which frequently remains active only for short periods, rotating domains, scripts, and delivery mechanisms to avoid sustained exposure. The fact that this behavior is short-lived reduces the effectiveness of retrospective analysis and limits the usefulness of static threat intelligence, requiring detection systems to

operate under strict time constraints. To combat this type of cybercrime, certain detection strategies include rule-based approaches, which rely on predefined indicators, and behavioral analysis using machine learning to classify URLs based on patterns and features. Lexical and web-scraped features, such as URL length, domain age, WHOIS information, and HTML structure, improve classification accuracy when combined (Nowroozi et al., 2022). Another challenge arises from the differences between detection accuracy and operational scalability, which make it so that the digital platforms that are already processing vast volumes of advertisements and user interactions in real time leave limited computational resources for deep inspection or complex behavioral analysis. As a result, security mechanisms must balance precision with performance, often prioritizing lightweight detection methods that may be less effective against sophisticated or well-obfuscated malvertising campaigns. Malvertising campaigns base the mode of operation on the exploitation of social trust and branding cues to increase perceived legitimacy, particularly through counterfeit or fake advertisements and popular styles of posts that mimic sponsorships (Grigsby, 2020). Additionally, rehosting vulnerabilities allow attackers to bypass protections, inject scripts, and persist in client-side attacks, reflecting the socio-technical complexity of malvertising as both a lure and a payload delivery mechanism (Watanabe et al., 2020). The involvement of multiple intermediaries and third-party services obscures the origin of malicious content, making it difficult to trace attacks back to their source, making it challenging to attribute faults and respond to incidents. Finally, getting to the main problem is that the lack of transparency and accountability makes clear that the takedown efforts and delays in responses are allowing campaigns to persist across different platforms and regions. One way to reduce the attacks of malvertising exploits is to install ad and tracking blockers to mitigate exposure by preventing ad and tracker requests (Borgolte & Feamster, 2020). Tackling these challenges together demonstrates that malvertising detection cannot rely solely on isolated technical indicators or user-side protections. Instead, effective defense requires adaptive mechanisms that account for the dynamic behavior of advertising content, the scale of modern ad delivery systems, and the human factors that influence user interaction. Addressing these dimensions collectively is essential for improving resilience against evolving malvertising threats.

3. Methodology

The study structure selected for this scientific literature on malvertising as a vector of cybercrime in digital platforms was the bibliometric approach. Data was retrieved from The Lens database, selected for its extensive coverage, detailed metadata exports, and compatibility with bibliometric tools.

The PRISMA 2020 framework (Figure 1) guided the selection process, providing a structured approach to ensure transparency and rigor in paper screening. With the initial search returning 442 records. Filters for document type (journal article), presence of abstract and full-text, and English language, along with duplicate removal, were applied, resulting in a final dataset of 236 papers.

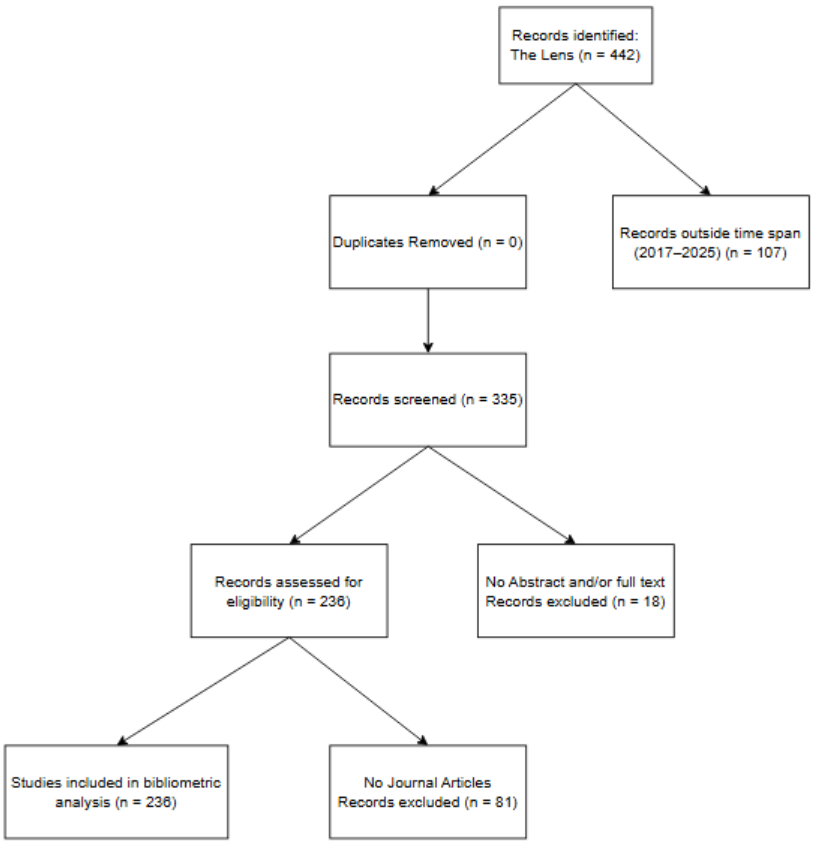

Fig 1. PRISMA flow diagram
Source: Authors' elaboration.

The search query was created to capture the scientific literature related to malicious advertising as a cybersecurity threat. Recurring to the usage of a combination of multiple terms related to malvertising, including malicious advertising, ad-based malware, ad fraud, drive-by downloads, exploit kits, and malware distribution through online advertisements,

ensuring broad coverage of attack techniques and delivery mechanisms. Then these terms are restricted by cybersecurity-related concepts, such as cybersecurity, information security, web security, browser exploits, malware delivery, and threat vectors, to focus the results on security-oriented studies. The query also includes platform-level and ecosystem-specific terms, in particular advertising platforms, ad networks, ad exchanges, adtech, and real-time bidding (RTB), enabling the identification of research addressing the technical and infra-structural dimensions of malvertising within digital advertising environments. With this structured combination, the aim is to have both thematic relevance and comprehensive cov-erage of the domain.

In the metadata extracted, it included authors, publication years, keywords, affiliations, and sources. The analyses were conducted using VOSviewer and Bibliometrix (RStudio) to gen-erate co-authorship networks, identify clusters of collaborating researchers, and map key-word co-occurrence to detect prevalent themes and subtopics. Density and topic maps were used to visualize thematic clusters, and annual publication illustrated the evolution of re-search activity and influence over time. Additional metrics taken included the identification of the most relevant sources, providing an overview of collaboration patterns and thematic structures in malvertising research.

Metadata	Description	Missing Counts	Missing %	Status
DI	DOI	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
SO	Journal	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
AB	Abstract	2	0.85	Good
AU	Author	2	0.85	Good
DE	Keywords	43	18.22	Acceptable
ID	Keywords Plus	43	18.22	Acceptable
CR	Cited References	54	22.88	Poor
C1	Affiliation	236	100.00	Completely missing
RP	Corresponding Author	236	100.00	Completely missing
LA	Language	236	100.00	Completely missing
WC	Science Categories	236	100.00	Completely missing

Fig 2. Density visualization of the metadata
Source: Bibliometrix output (2025).

4. Results

This segment offers a detailed descriptive overview of the bibliometric outcomes, establishing the basis for the following discussion and examination of the results. The findings are illustrated through network maps, tables, and visual analyses, providing an overview of the scientific research done about the selected literature. The results are organized to highlight key structural components of the research domain, such as publication trends, author collaboration networks, and the most prevalent themes and concepts. Each figure and table is examined individually, ensuring a clear and structured comprehension of the identified trends.

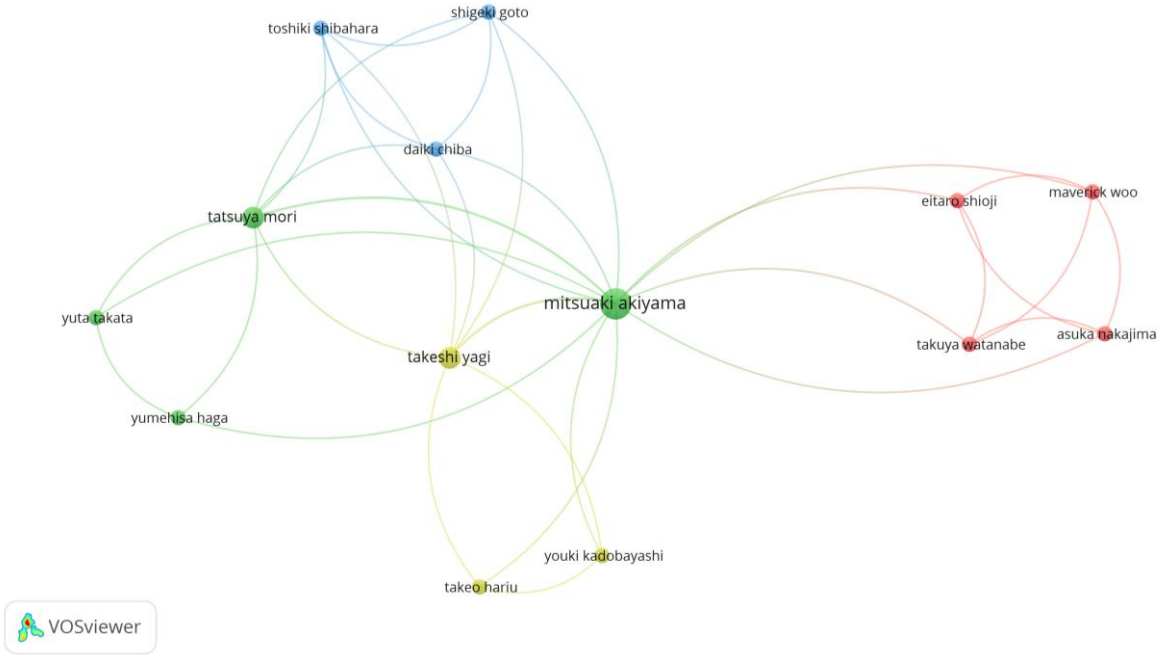

Fig 3. Co-authorship network of authors
Source: VOSviewer outputs (2025).

The co-authorship network (Figure 3) displays a map of the information collected from The Lens database and created with VOSviewer, highlighting the collaborative framework of scientific inquiry into malvertising as a means of cybercrime on digital platforms. The network comprises authors who have published at least one document and possess no fewer than two co-authorship links, guaranteeing that the map reflects recognized collaboration patterns instead of individual contributions. Authors are categorized into unique clusters, with each cluster representing research teams or collaborative groups created through co-authorship of publications.

It also illustrates various clusters that depict a fragmented but organized research environment. Writers with greater total link strength and citation totals hold more central roles in their clusters, signifying a stronger collaborative influence and greater scientific impact. Also, the smaller clusters are made up of authors who have fewer works and lower citation metrics, indicating more niche or developing research areas. The physical distance among clusters signifies restricted collaboration between groups, representing chances for wider cooperation among research communities to enhance and unify knowledge generation in the realms of malvertising and cybersecurity.

The network illustrated in Figure 4 represents the co-occurrence of author-defined keywords, considering a minimum occurrence threshold of five documents. This visualization captures the underlying conceptual organization of the literature on malvertising and cybersecurity, revealing how frequently used terms are interconnected across the analyzed studies. Central nodes such as “cybersecurity”, “cyber security”, and “cybercrime” occupy important positions in the network, reflecting their transversal role across the analyzed publications. These terms show strong connections with keywords related to malicious activities and threat vectors, including “malware”, “ransomware”, “phishing”, and “cyber attacks”, indicating a research focus on the use of online platforms and advertising infrastructures as key objectives for cybercrime.

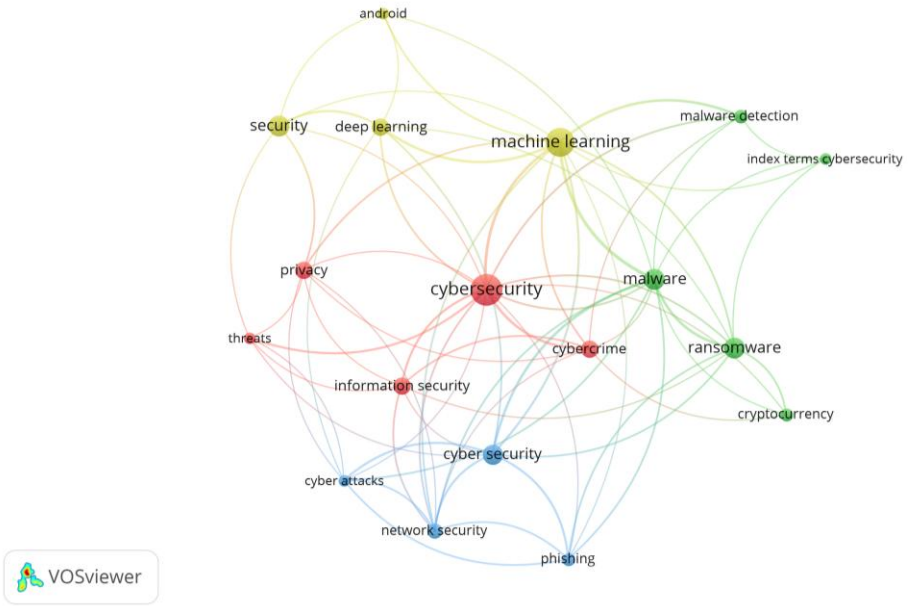

Fig 4. Keyword co-occurrence network
Source: VOSviewer outputs (2025).

Distinct thematic clusters can be observed. One cluster emphasizes technical detection and mitigation approaches, bringing together terms such as “machine learning”, “deep learning”, and “malware detection”, which suggests a growing reliance on data-driven techniques to identify malicious advertising campaigns. Another cluster is associated with platform and ecosystem contexts, including keywords such as “android”, “network security”, and “privacy”, highlighting concerns related to mobile environments and user protection. Overall, the keyword co-occurrence network illustrates a field structured around the intersection of cybersecurity, malware distribution mechanisms, and advanced analytical techniques, with increasing attention to automated detection and platform-specific threats.

Figure 5 represents the co-occurrence network of frequently used terms taken from the reviewed papers, focusing solely on terms referenced at least ten times and appearing in multiple articles. This map emphasizes frequently employed concepts in literature, providing a comprehensive summary of the prevalent terminology and recurring research topics in investigations related to malvertising and cybersecurity.

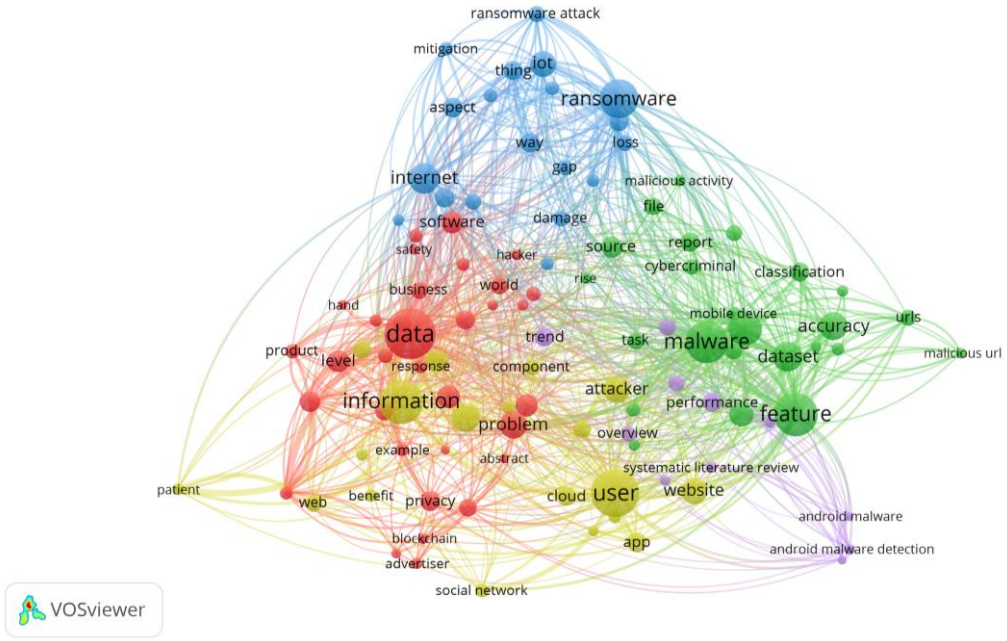

Fig 5. Common-term co-occurrence network in malvertising and cybersecurity literature
Source: VOSviewer outputs (2025).

The network shows multiple dense clusters that indicate unique yet interrelated research domains. A significant grouping highlights ideas related to data and detection, including terms like data, algorithm, feature, dataset, machine learning, and malware detection, reflecting a strong focus on analytical and computational methods. A different cluster organizes terms associated with threats and attack dynamics, including attacker, cybercrime,

ransomware, botnet, phishing, and malicious activities, emphasizing the operational viewpoint of cybercrime activities. The topics mobile and platforms arise from recurring terms such as android, mobile device, app, online advertising, and websites, highlighting the significance of digital platforms as avenues for harmful actions. The occurrence and concentration of terms indicate that the literature merges particular detection techniques with wider discussions regarding cyber threats, user impacts, and platform vulnerabilities, underscoring the interdisciplinary aspect of malvertising studies.

The evolution of scientific output from 2017 until 2025 shows significant annual variations in publication counts (Figure 6). From 2017 to 2019, the total of publications rose from 21 to 32, reflecting an increasing focus on research in this area. Scientific output reached its highest point in 2020 with 48 publications, then experienced a minor drop to 40 in 2021, before a small increase to 44 in 2022. Following 2022, a significant decline is observed, leading to 21 publications in 2023 and only seven in each of 2024 and 2025, mainly due to constraints in data gathering for the recent years. This trend illustrates the dynamic nature of research in malvertising and cybersecurity, highlighting periods of intensified academic attention as well as the current stabilization phase.

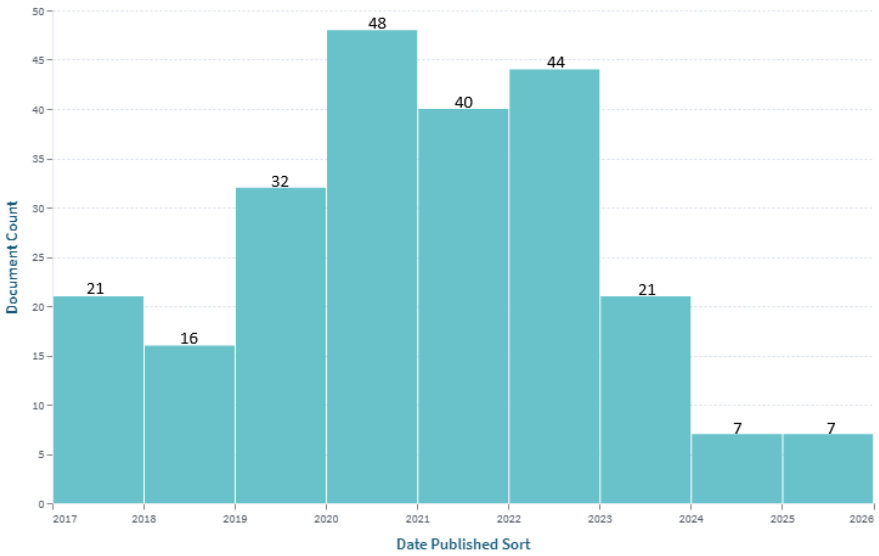

Fig 6. Annual scientific production on malvertising-related cybersecurity research
Source: The Lens database (2025).

The thematic map consists of four primary quadrants (Figure 7): Motor Themes, Basic Themes, Niche Themes, and Emerging or Declining Themes, arranged according to two axes: development degree represented on the vertical axis and relevance degree included on the horizontal axis. Fields such as machine learning, cybersecurity, and deep learning are

well developed and prominent in the upper-right quadrant with the Motor Themes, signifying established domains with strategic significance for upcoming research. In the lower-right quadrant are included the Basic Themes with malware, ransomware, cryptocurrency, cybersecurity, and information security, which constitute the foundation of the field. While not as focused as motor themes, these topics show considerable importance but lower density, providing a conceptual and practical foundation for study.

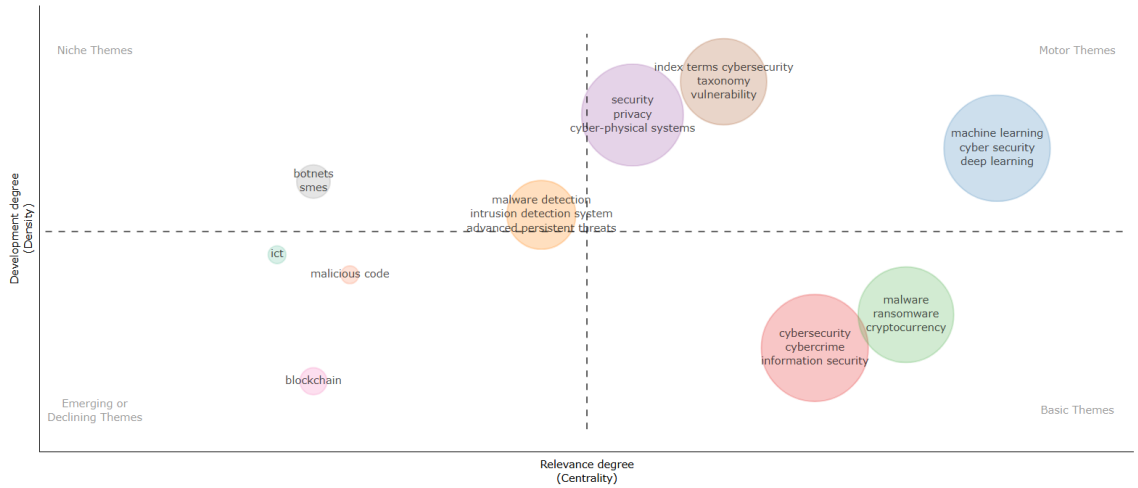

Fig 7. Thematic map of research topics on malvertising and cybersecurity
Source: Bibliometrix (2025).

The upper-left quadrant represents the Niche Themes with subjects such as security privacy, cyber-physical systems, and vulnerability taxonomy. These subjects show high density but low centrality, suggesting extensive research in specific subfields that have minimal broader influence. The lower-left quadrant appears the Emerging or Declining Themes which feature Blockchain, ICT, and malicious code, characterized by low density and low centrality, signifying either initial development or possible decline. In conclusion, this map illustrates that the incorporation of artificial intelligence in cybersecurity, particularly via machine learning and deep learning, is presently guiding research initiatives, while malware and ransomware remain crucial topics. With the continuous advancement of technology and cyber threats, emerging fields like blockchain and malicious software may gain greater importance.

Table 1 shows the most relevant sources in scientific production in malvertising and cybersecurity. The data indicate that a sizable portion of the papers is concentrated in leading sources, such as IEEE Access (17 articles), Sensors (Basel, Switzerland) (13 papers), and

Electronics (12 papers). Other important sources include Applied Sciences (8 papers), SSRN Electronic Journal (6 papers), Security and Communication Networks (5 papers), Information (4 papers), International Journal of Advanced Computer Science and Applications (4 papers), International Journal of Information Security (4 papers), and Journal of Cyber Security and Mobility (4 papers). These results highlight the sources that contribute most to the field, reflecting the thematic spread across cybersecurity, computer science, and electronic engineering. Also, let's not forget that even if these sources do fit in with the top publication sources, they only compose 77 papers out of the 236 analyzed. This means that smaller sources also impacted the field with research, with 159 papers published, but some with only 2/3 maximum per source.

Table 1. Top 10 publication sources in malvertising and cybersecurity research

Source	Articles
IEEE Access	17
Sensors (Basel, Switzerland)	13
Electronics	12
Applied Sciences	8
SSRN Electronic Journal	6
Security and Communication Networks	5
Information	4
International Journal of Advanced Computer Science and Applications	4
International Journal of Information Security	4
Journal of Cyber Security and Mobility	4

Source: Authors' elaboration based on data from The Lens database (2025).

5. Discussion

5.1. Patterns and Trends

The bibliometric review reveals that conventional blacklist approaches for detection are inadequate to manage the changing and evolving characteristics of harmful URLs, stressing the need for adaptive and smart detection methods. The significance of regularly refreshing detection models and refining feature selection is highlighted, guaranteeing that new attack patterns can be successfully addressed within digital advertising networks (Shin et al., 2022). Exploit kits paired with malvertising demonstrate the weaponization of advertising

ecosystems, facilitating widespread infection by diverting users from legitimate ad spaces to compromised landing pages (Usharani et al., 2021).

5.2. Analysis of Key Authors, Institutions, and Sources

The leading sources and authors indicate a focus of research activities in influential publications and niche investigations. Although bibliometric data does not enable a direct analysis at the country's level, the significance of institutions like IEEE Access, Sensors, and Electronics (leading sources) suggests a concentration of academic activity in technologically advanced research centers. The networks of co-authorship and patterns of institutional production indicate a cooperative atmosphere, frequently among authors who emphasize detection methods and analyses at the ecosystem level. These results, aligned with earlier studies, highlight the importance of centralized research teams in enhancing understanding of malvertising and cybersecurity.

Comparison with Previous Studies

Previous studies show that malicious ads often give themselves away through long URLs, fake domains, or direct IP addresses. However, the GandCrab project proved that malvertising has evolved into much more than simple ad fraud. It is now a primary way to spread ransomware, which creates an urgent need for better security in ad networks and more advanced tools to catch exploit kits (Nowroozi et al., 2022). At the same time, bot-driven spam remains a major risk because these automated systems are now excellent at mimicking real human behavior (Mbona & Eloff, 2021). More recently, research has found that malvertising is becoming a standard part of the Ransomware-as-a-Service model. This trend exploits the vulnerabilities found in automated bidding and the industry's heavy reliance on third parties (Anand et al., 2023). Overall, these findings show that we should no longer view malvertising as a series of isolated tricks but rather as a systemic threat to the entire digital ecosystem.

Thematic Analysis and Evolution of Topics

The thematic clusters identified in the keyword analysis reveal an evolution from intrusion detection-focused research to approaches emphasizing machine learning, behavioral analytics, and ecosystem resilience. Malvertising's systemic impact includes user compromise, erosion of ecosystem trust, and financial losses, often compounded by click fraud to

maximize revenue. The discussion of countermeasures includes consistent validation of ad content, the implementation of anti-malware tools, and detection techniques based on machine learning, like SVM classifiers trained on features of suspicious ads. Nonetheless, current models are usually confined to recognized attack patterns, rendering systems susceptible to zero-day malvertising risks. Newly developed solutions, such as game-theoretic defenses, dynamic traffic analysis, and blockchain-driven transparency, offer ways to strengthen ad-tech security principles and address the multifaceted, programmatic, and RTB-enhanced dangers brought by malvertising (Pooranian et al., 2021). Significantly, sophisticated methods like PhishTransformer utilize CNN and Transformer encoders to examine URLs in ad frames and scripts, attaining cutting-edge precision, recall, and F1-scores while simplifying detection complexity (Asiri et al., 2024). Moreover, machine learning and data-driven methods, such as publisher-side SVM classifiers and MadTracer rules, facilitate extensive malvertising detection, although they depend significantly on current heuristics and strong feature engineering. The literature further anticipates that emerging platforms such as IoT and AR/VR will become future malvertising targets, highlighting the need to apply current defense mechanisms, such as sandboxing, behavioral analysis, and secure ad SDKs, to these new digital endpoints (Chua et al., 2020).

6. Conclusion

This bibliometric analysis offers an extensive overview of malvertising research, highlighting trends in publication, author partnerships, and thematic development. The findings underscore the shortcomings of conventional detection methods, like blacklists, and stress the rising use of adaptive, machine learning-driven strategies to combat advancing cyber threats. The co-authorship networks indicate possibilities for wider interdisciplinary cooperation, whereas leading sources and recognized research groups prevail in the domain. This research focuses on The Lens database and English-language publications, revealing prominent research domains such as malware detection, ecosystem vulnerabilities, and AI-based defense methods. In addition, new topics like blockchain incorporation, IoT and AR/VR security, and zero-day malvertising, presenting possibilities for future research.

In conclusion, the information available shows that malvertising serves not just as a means of ad fraud but also as a pervasive cybercrime risking users, platforms, and advertising

networks. The findings offer important perspectives for both researchers and users, informing the creation of stronger detection systems, real-time surveillance frameworks, and forward-thinking cybersecurity strategies. With the evolution of malvertising, the future of research will be based on adaptive defense mechanisms, evolving digital platforms, and cross-disciplinary methods required to address the increasing issues caused by malvertising, using integrated temporal, collaborative, and thematic aspects of research.

References

Anand, P. M., Charan, P. V. S., & Shukla, S. K. (2023). HiPeR – Early detection of a ransomware attack using hardware performance counters. *Digital Threats: Research and Practice, 4*(3), Article 43, 1–24. https://doi.org/10.1145/3608484

Arrate, A., González-Cabañas, J., Cuevas, Á., & Cuevas, R. (2020). Malvertising in Facebook: Analysis, quantification and solution. *Electronics, 9*(8), 1332. https://doi.org/10.3390/electronics9081332

Asiri, S., Xiao, Y., & Li, T. (2024). PhishTransformer: A novel approach to detect phishing attacks using URL collection and transformer. *Electronics, 13*(1), 30. https://doi.org/10.3390/electronics13010030

Borgolte, K., & Feamster, N. (2020). Understanding the performance costs and benefits of privacy-focused browser extensions. In *Proceedings of The Web Conference 2020* (pp. 2275–2286). ACM. https://doi.org/10.1145/3366423.3380292

Chua, M. Y.-K., Yee, G. O. M., Gu, Y. X., & Lung, C.-H. (2020). Threats to online advertising and countermeasures: A technical survey. *Digital Threats: Research and Practice, 1*(2), Article 11, 1–27. https://doi.org/10.1145/3374136

Caviglione, L., Choras, M., Corona, I., Janicki, A., Mazurczyk, W., Pawlicki, M., & Wasielewska, K. (2021). Tight arms race: Overview of current malware threats and trends in their detection. *IEEE Access: Practical Innovations, Open Solutions*, *9*, 5371–5396. https://doi.org/10.1109/access.2020.3048319

Grigsby, J. L. (2020). Fake ads: The influence of counterfeit native ads on brands and consumers. *Journal of Promotion Management*, *26*(4), 569–592. https://doi.org/10.1080/10496491.2020.1719958

Kasturi, R. P., Fuller, J., Sun, Y., Chabklo, O., Rodriguez, A., Park, J., & Saltaformaggio, B. (2022). Mistrust plugins you must: A large-scale study of malicious plugins in WordPress marketplaces. In *Proceedings of the 31st USENIX Security Symposium* (pp. 161–178). USENIX Association.

Liu, T., Wang, H., Li, L., Luo, X., Dong, F., Guo, Y., Wang, L., Bissyandé, T. F., & Klein, J. (2020). MadDroid: Characterizing and detecting devious ad contents for Android apps. In

Proceedings of The Web Conference 2020 (pp. 1715–1726). ACM. https://doi.org/10.1145/3366423.3380242

Mbona, I., & Eloff, J. H. P. (2022). Feature selection using Benford's law to support detection of malicious social media bots. *Information Sciences, 582,* 369–381. https://doi.org/10.1016/j.ins.2021.09.038

Nowroozi, E., Abhishek, Mohammadi, M. R., & Conti, M. (2022). An adversarial attack analysis on malicious advertisement URL detection framework. *IEEE Transactions on Network and Service Management.* Advance online publication. https://doi.org/10.48550/arXiv.2204.13172

Pooranian, Z., Conti, M., Haddadi, H., & Tafazolli, R. (2021). Online advertising security: Issues, taxonomy, and future directions. *IEEE Communications Surveys & Tutorials, 23*(4), 2494–2524. https://doi.org/10.1109/COMST.2021.3118271

Sadeghpour, S., & Vlajic, N. (2021). Ads and fraud: A comprehensive survey of fraud in online advertising. *Journal of Cybersecurity and Privacy, 1*(4), 804–832. https://doi.org/10.3390/jcp1040039

Shin, S.-S., Ji, S.-G., & Hong, S.-S. (2022). A heterogeneous machine learning ensemble framework for malicious webpage detection. *Applied Sciences, 12*(23), 12070. https://doi.org/10.3390/app122312070

Usharani, S., Manju Bala, P., & Martina Jose Mary, M. (2021). Dynamic analysis on crypto-ransomware by using machine learning: GandCrab ransomware. *Journal of Physics: Conference Series, 1717,* 012024. https://doi.org/10.1088/1742-6596/1717/1/012024

Watanabe, T., Shioji, E., Akiyama, M., & Mori, T. (2020). Melting pot of origins: Compromising the intermediary web services that rehost websites. *Proceedings of the Network and Distributed Systems Security (NDSS) Symposium 2020* (pp. 1–15). Internet Society. https://doi.org/10.14722/ndss.2020.24140

https://doi.org/10.58086/t0n3-0y36

SECURITY AND RISK IN SOFTWARE DEVELOPMENT PROJECTS: A BIBLIOMETRIC REVIEW

Francisco Conceição[1]✉, **Mário Dias Lousã**[1,2], **José Carlos Morais**[1,3]

[1] Instituto Superior Politécnico Gaya (ISPGAYA), Portugal.
[2] Insight - Piaget Research Center for Ecological Human Developement, Portugal.
[3] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal.
✉ Corresponding authors**:** ispg2024100498@ispgaya.pt

Abstract

Security analysis is increasingly central to software development as organizations face rising cyber risk and regulatory pressure. Although extensive research exists on cyber risk assessment, secure software development, and security requirements, the literature remains fragmented at the project level. This study presents a bibliometric analysis of research published between 2015 and 2026, using data retrieved exclusively from The Lens and structured through a PRISMA-guided workflow. Only journal and conference publications addressing security analysis within software development projects were retained, while studies focused solely on isolated technical vulnerabilities or non-project contexts were excluded, resulting in a final dataset of 1,008 documents. The dataset was analyzed using descriptive bibliometrics, collaboration and geographical analysis, field-of-study classification, keyword co-occurrence, co-citation, and bibliographic coupling, supported by VOSviewer. Results show sustained growth after 2019 and strong dominance of computer science and software engineering. Influential contributions cluster around secure SDLC frameworks, ISO/NIST standards, requirement decomposition, and emerging quantitative risk models. Despite this consolidation, the analysis reveals persistent gaps, including weak integration between cyber risk assessment and requirements engineering, limited project-level operationalization of security attributes, and a scarcity of approaches tailored to small and medium-sized enterprises (SMEs). These findings highlight the need for integrated, requirement-driven security analysis frameworks that bridge technical and organizational perspectives within software development projects.

Keywords: SMEs, Software projects, Cyber risk assessment, Requirements engineering, Bibliometric Analysis.

1. Introduction

SMEs are key drivers of economic growth and innovation but remain highly vulnerable to cyber threats. Many lack the financial resources, cybersecurity expertise, and organizational maturity needed to implement robust security and risk-management practices (Brown & Thorpe, 2023). As operations increasingly rely on internally developed software, new projects introduce significant risk when integration points, dependencies, and data flows are not systematically assessed. These pathways may create new attack surfaces, alter trust boundaries, or expose sensitive information, making project-level risk assessment a critical yet often neglected aspect of SME cybersecurity governance.

Academic interest in cyber risk assessment, secure software development life cycles (SDLC), requirements engineering, and security-aware requirement decomposition have expanded, yet research remains fragmented, particularly in SME-focused studies. Evidence shows that many SMEs underestimate cyber threats and deprioritize risk management due to limited resources and cybersecurity literacy (Rizvi et al., 2023). This has produced a persistent gap between high-level governance frameworks and the operational realities of SMEs developing software under technical, organizational, and regulatory constraints.

No existing framework integrates three elements in a manner tailored to SMEs: systematic requirement decomposition, structured mapping of security attributes such as confidentiality, integrity, and availability (CIA), and quantitative project-level risk scoring, including in regulatory contexts such as the EU Cyber Resilience Act (CRA). Current approaches address only fragments of this problem. SDLC standards encourage early security integration but offer little guidance for requirement-level analysis. Requirements engineering models provide decomposition structures but rarely incorporate security attributes systematically. Quantitative risk methods often operate at the organizational level or demand resources unrealistic for SMEs. Although some studies propose lightweight or SME-oriented approaches (El-Hajj et al., 2024), these typically lack requirement-level structuring and rigorous CIA-based analysis. As a result, SMEs lack systematic tools to evaluate the cyber-risk impact of new software components before integration.

These limitations create a clear need to examine how research in this domain has evolved, including which frameworks dominate, how risk assessment and requirements engineering interact, and which gaps persist.

This study responds through a bibliometric analysis of literature published between 2015 and 2026. It charts thematic evolution, identifies influential frameworks and research communities, visualizes conceptual clusters (secure SDLC, requirement decomposition, CIA mapping, and SME governance), and highlights structural gaps motivating the development of an integrated cyber-risk assessment framework for internal SME software projects. By quantifying scientific output and mapping the field's intellectual structure, the study contributes to academic understanding and supports the development of practically adoptable risk-assessment approaches.

2. Key Concepts and Terms

This section introduces and clarifies the core concepts and terms that underpin the bibliometric analysis presented in this paper. While these concepts are now widely used, their interpretation and operationalization have evolved over time and vary across different research traditions.

2.1 Security Analysis

Security analysis refers to the systematic evaluation of security concerns across the software lifecycle, including requirements, design, implementation, and governance. Earlier approaches treated security as a late technical activity, whereas recent research frames it as a methodological and decision-oriented process shaping architecture and project outcomes (Assal & Chiasson, 2019; Granata, 2024).

2.2 Cyber Risk Assessment

Cyber risk assessment involves identifying and evaluating threats, vulnerabilities, and impacts affecting software systems and development processes. Contemporary studies emphasize design-time and requirements-level analysis, as early decisions strongly influence exposure and mitigation costs (El Amin et al., 2024; Radanliev et al., 2020).

2.3 Requirements Engineering

Requirements engineering concerns the elicitation, analysis, specification, and management of project requirements and is increasingly viewed as a key entry point for integrating security (Quiñones et al., 2018). Existing approaches include goal-oriented, threat-driven, risk-based, and compliance-focused perspectives, contributing to conceptual fragmentation.

2.4 Secure Software Development Lifecycle (SDLC)

Secure SDLC approaches integrate security activities across development phases, supporting earlier risk identification, improved team coordination, and resilience to supply chain threats (McGraw, 2019; Myrbakken & Colomo-Palacios, 2017).

3. Research Questions

This study is guided by a set of research questions designed to map, structure, and interpret the scientific landscape surrounding security analysis within software development.

RQ1: How has the scientific production on cyber risk assessment for software development evolved over time?

Examines how scientific production on cyber risk assessment for software development has evolved over time. Analyzing publication trends provides Insight into the maturation of the field, the emergence of methodological turning points and the influence of regulatory or technological shifts, allowing assessment of whether the domain is expanding, stabilizing or fragmenting.

RQ2: Which authors and institutions have most influenced this research domain?

Investigates which authors and institutions have exerted the greatest influence on this research domain. By analyzing authorship patterns, collaboration networks, and institutional concentration, the study identifies intellectual centers, assesses the coherence of the research community, and explores how academic and industrial environments shape methodological development.

RQ3: Which countries contribute most to research on cyber risk assessment in software development?

Focuses on the geographical distribution of scientific output, identifying the countries that contribute most to research on cyber risk assessment in software development. This analysis highlights regional asymmetries, centers of scientific activity, and the influence of national contexts, while also assessing the field's degree of internationalization.

RQ4: Which cyber risk assessment frameworks, requirement decomposition models, and disciplinary domains dominate the literature on security analysis in software development?

Examine the dominant cyber risk assessment frameworks, requirement decomposition models, and disciplinary domains underpinning literature. It explores how security analysis is framed across academic fields and the extent to which existing approaches integrate or separate risk analysis, requirements engineering, and security attribute mapping.

RQ5: How do keywords cluster conceptually to reveal core themes?

Analyze keyword clustering to reveal core conceptual themes and their interrelationships.

4. Bibliometric Methodology

This study applies the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Figure 1) guidelines to ensure transparency and replicability. While PRISMA was designed for systematic reviews, its framework effectively documents identification, screening, eligibility, and inclusion in bibliometric research. The methodology also incorporates co-citation, bibliographic coupling, and keyword co-occurrence analyses.

4.1 Databases Used

The bibliographic dataset was retrieved exclusively from The Lens, which aggregates metadata from major sources such as CrossRef, PubMed and Microsoft Academic Graph and provides standardized records suitable for bibliometric analysis. Using a single source ensured consistency, reduced duplication, and preserved comparability across retrieved records.

The bibliographic search was conducted in The Lens on 20 December 2025. Bibliometric network construction and visualization were performed using VOSviewer (version 1.6.20), which supported co-authorship, co-citation, bibliographic coupling, and keyword co-occurrence analyses. Data cleaning, filtering, and descriptive statistical analysis were carried out prior to network generation to ensure the consistency and reliability of the bibliographic records.

4.2 Dynamic Anonymization Algorithm

The search strategy was developed iteratively to balance conceptual breadth and thematic precision. Exploratory searches were used to refine terminology and boundaries and to ensure coverage of research on security analysis in software projects. The final query reflects the convergence of three domains: cyber-risk assessment, security-aware requirements engineering and software development life cycles, and was applied in The Lens to retrieve publications framing security as a methodological component of software projects:

(("cyber risk assessment" AND "software development") OR ("software requirements" AND "security" AND "risk") OR ("risk assessment" AND "software project"))

To reflect contemporary developments in cybersecurity governance, secure software engineering, and regulations such as NIS2 and the CRA, the search was limited to publications from 2015 to 2026. This timeframe captures both the early emergence of requirement-level security models and the later shift toward integrated risk-aware development practices. All records were exported in structured bibliographic format for screening and refinement through the PRISMA workflow.

4.3 PRISMA Workflow - Inclusion/Exclusion Criteria

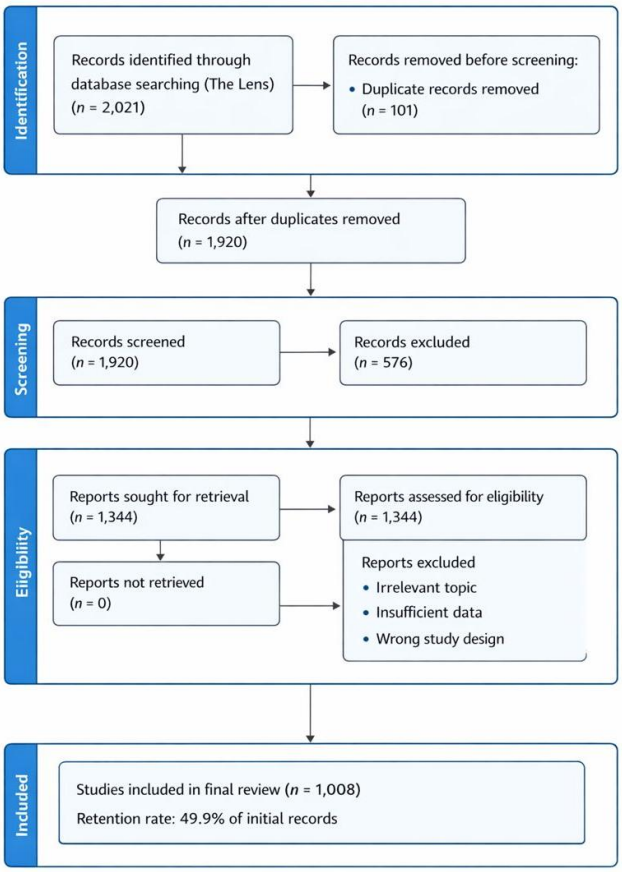

Fig 1. PRISMA flow diagram illustrating the study selection process
Source: Authors' elaboration based on PRISMA statement (no date).

Identification

The identification stage aimed to capture the broadest possible set of publications relevant to the security analysis of software projects. A comprehensive search was conducted exclusively in The Lens using the final, refined query designed to capture the intersection of cyber risk assessment, security-oriented requirements engineering, and software project contexts.

This search returned 2,021 records, representing a wide range of disciplinary perspectives. Prior to screening, 101 duplicate records were identified and removed based on metadata comparison, resulting in 1,920 unique records entering the screening phase.

Screening

The screening stage focused on eliminating records that were clearly outside the scope of the study without requiring detailed conceptual analysis. Titles and abstracts of the 1,920 records were reviewed to assess basic relevance and completeness. During this phase, 576

records were excluded because they did not address security analysis in software development contexts, lacked sufficient descriptive information, or fell outside academic literature.

Technical reports, editorial notes, white papers, and unreviewed preprints were also excluded at this stage. The screening process ensured that the remaining corpus was both methodologically credible and thematically aligned with the study's objectives, leaving 1,344 records.

Eligibility

The eligibility stage assessed the 1,344 publications that passed screening by reviewing abstracts, keywords, and full texts to determine substantive contributions to software-project security analysis. Studies were considered eligible if they proposed or applied frameworks, models, or methods supporting structured security evaluation within software development life cycles.

Publications focusing mainly on isolated vulnerabilities, penetration testing, or narrow threat enumeration, as well as those limited to high-level organizational governance without project-level methodology, were excluded. As a result, 336 studies were removed due to thematic irrelevance, insufficient methodological contribution, or unsuitable design.

Inclusion

The final dataset comprises 1,008 publications, corresponding to a retention rate of 49.9% of the initially identified records. These studies form the empirical basis for all subsequent bibliometric analyses.

This curated dataset comprises publications that align security evaluation with software engineering processes, requirement decomposition, architectural considerations, and project-level governance. It provides an empirical foundation for the analyses presented in the subsequent sections. The complete progression from identification to inclusion is visually summarized in the PRISMA flow diagram included in this paper.

4.4 Analysis Procedures

The analytical phase characterized research on security analysis in software development using descriptive and relational bibliometric techniques applied to 1,008 publications extracted and normalized from The Lens.

Temporal analysis examined publication trends (2015–2026), identifying growth, acceleration, and stabilization phases in relation to regulatory and methodological developments. Authorship analysis assessed productivity and collaboration patterns to reveal the field's intellectual structure and concentration of influence.

Disciplinary analysis evaluated the distribution of publications across fields to determine the dominance of technical versus organizational perspectives. Keyword co-occurrence, co-citation, and bibliographic coupling analyses identified dominant themes, intellectual linkages, and structural gaps, forming the empirical basis for the results.

5. Results

This section presents the empirical findings of bibliometric analysis. The results are organized around five research questions, each addressing a distinct aspect of the scholarly landscape on security analysis in software development. Beyond summarizing the evidence, the chapter highlights structural patterns, conceptual orientations, and persistent gaps within the literature.

5.1. RQ1: Evolution of Scientific Output Over Time

The temporal distribution of publications shows (Figure 2) a clear evolution in scientific output between 2015 and 2026. From 2015 to 2018, publication activity remained stable, suggesting an early consolidation phase characterized by consistent research interest and limited methodological diversification.

From 2019 onwards, a pronounced upward trend emerges, peaking between 2021 and 2022. This period represents the highest level of scientific output and reflects a phase of accelerated academic interest. The surge coincides with the wider adoption of SDLCs, growing awareness of software supply chain risks, and the increasing influence of cybersecurity regulation and governance frameworks. This concentration of publications indicates that security analysis in software projects gained strategic relevance within both academic and research communities.

After 2022, publication counts decline but remain above pre-2019 levels, indicating stabilization rather than reduced interest. Lower values for 2025 and 2026 should be interpreted cautiously due to indexing delays.

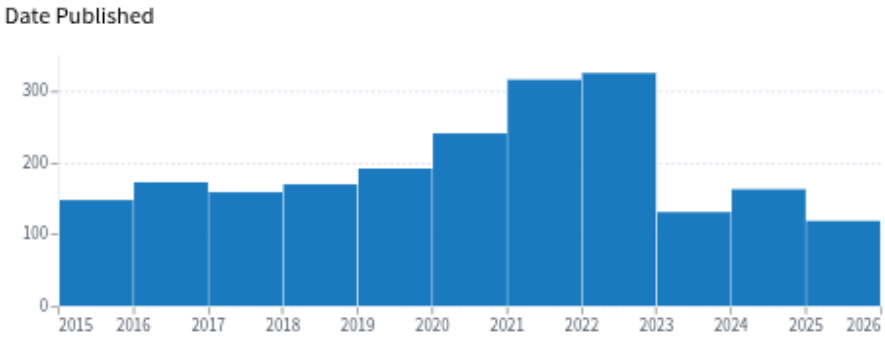

Fig 2. Scientific output on cyber risk assessment and security analysis in software development (2015–2026).
Source: The Lens database (2025).

5.2. RQ2: Influential Authors

The author profile analysis (Figure 2) reveals an uneven distribution of contributions, with a small group of researchers publishing repeatedly while most authors contribute only occasionally. This suggests that the field is influenced by limited core contributors rather than a consolidated research community.

Ruzanna Chitchyan emerges as the most prolific author, reflecting sustained engagement with requirements engineering and security-aware development. Alberto Rodrigues da Silva and Alok Mishra also show notable publication activity at the intersection of software engineering, project management, and security. Authors with two publications further reinforce the presence of limited continuity across the field.

The relatively low number of publications per author highlights the interdisciplinary and fragmented nature of the domain, with contributions often originating from adjacent research areas. This fragmentation aligns with broader bibliometric findings showing weak integration between cyber risk assessment and requirements engineering. Overall, intellectual leadership is present but not yet strongly centralized.

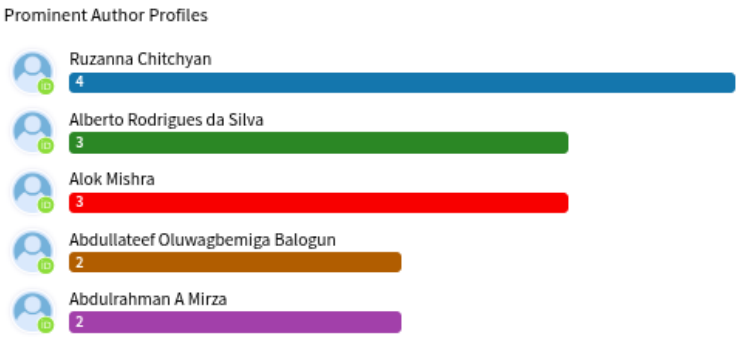

Fig 3. Prominent author profiles in research on cyber risk assessment and security analysis in software development (2015–2026).
Source: The Lens database (2025).

5.3. RQ 3: Geographical Distribution of Scientific Output

Scientific output is geographically concentrated in a small number of countries, producing an uneven global research landscape. Activity is strongest in North America, Western Europe, and parts of the Asia–Pacific region, with limited contributions from Africa, much of South America, and the Middle East (Figure 4).

The United States leads with 197 publications, reflecting strong academic infrastructure, sustained investment in cybersecurity, and close collaboration between academia, industry, and government. The United Kingdom follows with 134 publications, supported by a mature research ecosystem and a regulatory environment emphasizing cybersecurity governance. China contributes 111 publications, representing the largest output in the Asia–Pacific region and signaling a growing strategic emphasis on software security, primarily through technically oriented research.

Within Europe, Germany (72) demonstrates notable activity, often aligned with engineering-driven and standards-based approaches. Brazil (50) emerges as a key contributor in South America. India (61) and Australia (51) further reinforce Asia–Pacific representation.

Overall, dominant frameworks and methodologies are shaped mainly within technologically advanced economies.

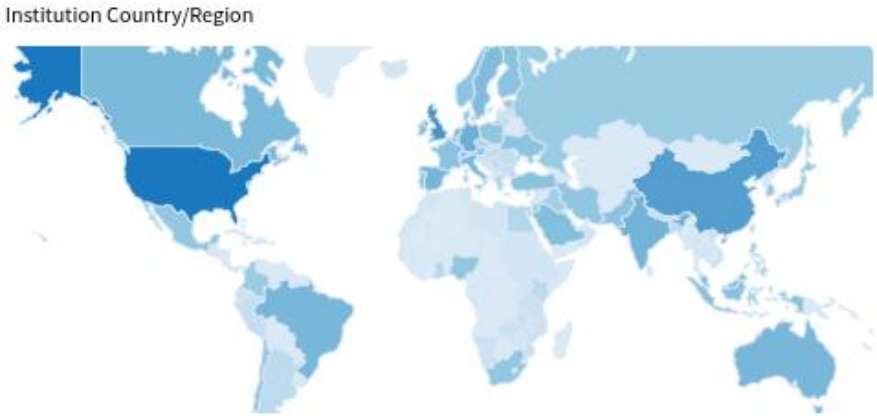

Fig 4. Global research output on cyber risk assessment in software development (2015–2026)
Source: The Lens database (2025).

5.4. RQ4: Dominant Frameworks, Conceptual Foundations and Disciplinary Origins

Literature converges around a limited set of influential frameworks. The NIST Secure Software Development Framework (SP 800-218) is a central reference, particularly in work addressing secure development practices and organizational governance. ISO/IEC 27034 and ISO/IEC 27005 also feature prominently, providing governance-oriented guidance and risk assessment structures often operationalized through threat–likelihood–impact models.

Within requirements engineering, decomposition-oriented approaches such as the Requirements Abstraction Model support analysis of how security requirements propagate across artefacts, and architectural layers, frequently structured around CIA. A smaller body of work explores quantitative risk assessment models inspired by FAIR, introducing probabilistic reasoning and scoring mechanisms that remain only partially integrated into software engineering practice.

To contextualize these frameworks, a field-of-study analysis was conducted. Computer Science dominates the dataset (Figure 5), framing security analysis primarily as a technical problem grounded in modelling, and computational reasoning. Software Engineering and Engineering provide secondary contributions, supporting secure SDLC practices and requirement decomposition, but largely retain a technical focus. Business and Information Systems fields contribute relatively little, suggesting underrepresentation of organizational and managerial perspectives and helping explain the scarcity of SME-oriented frameworks.

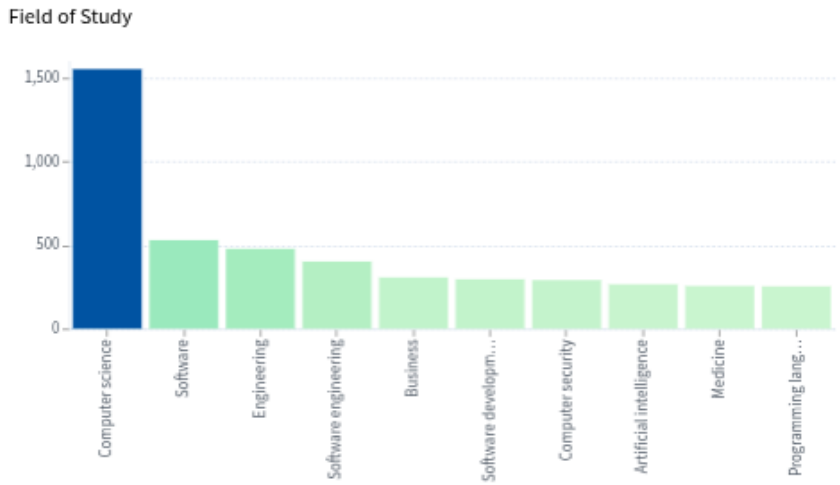

Fig 5. Dominant academic fields in security analysis research for software development projects (2015–2026).
Source: The Lens database (2025).

5.5. RQ5: Keyword Cluster Analysis

The keyword co-occurrence network (Figure 6) reveals a conceptually fragmented but moderately interconnected field. Distinct color clusters represent coherent thematic orientations, with link density indicating conceptual proximity and node position reflecting interaction across themes. Central nodes such as system, model, requirement, and risk act as bridges between clusters, while peripheral nodes represent specialized strands.

Red Cluster: System-Centered and Information-Oriented Security

The red cluster centers on system, security, data, information, application, and technology, reflecting a system-centric perspective focused on protecting information assets and architectures. Its strong internal connectivity suggests a mature research tradition rooted in computer science, though its separation from project- and requirement-oriented terms indicates limited integration with development processes.

Green Cluster: Project, Risk, and Development Lifecycle Orientation

The green cluster groups risk, project, software development, model, and technique, representing research that situates security within software projects and development life cycles. While it emphasizes decision-oriented and process-aware analysis, its limited overlap with system-level concepts suggests abstraction from concrete security attributes.

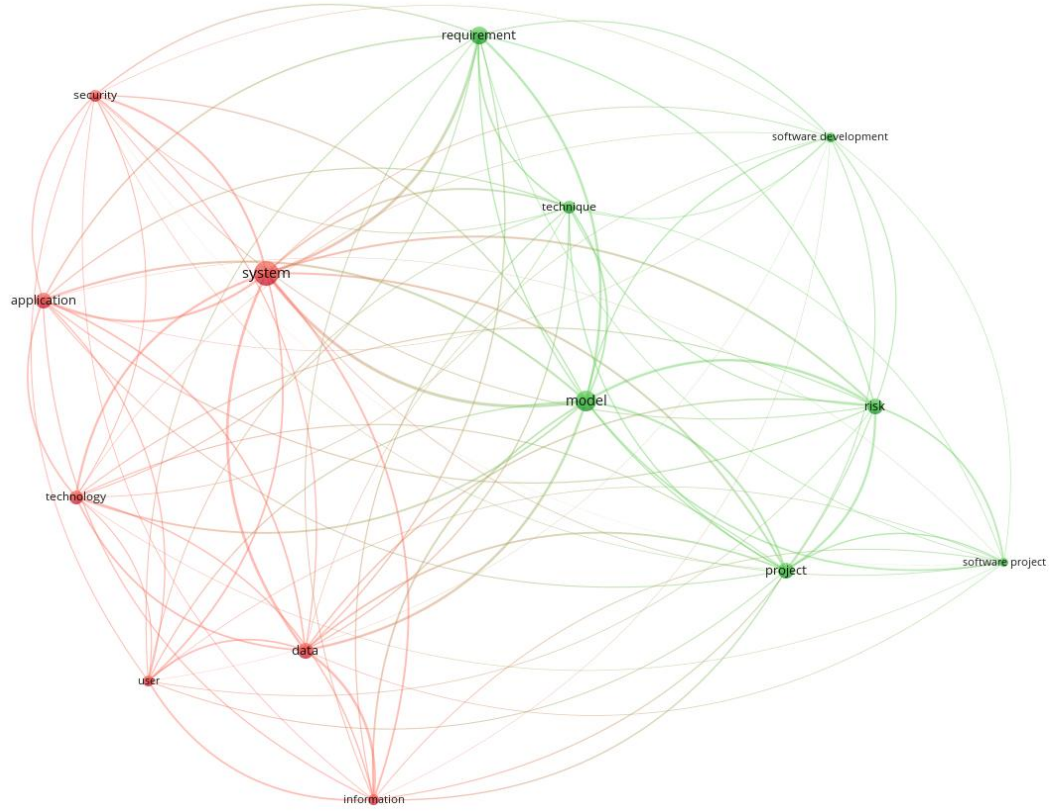

Fig 6.Conceptual structure of security analysis research in software development projects (2015–2026)
Source: VOSviewer (2025).

Light Green Cluster: Requirements and Methodological Abstraction

The light green cluster focuses on requirement, model, and technique, acting as a conceptual bridge between system and project perspectives. Although the requirement appears as a highly connected node, the relative sparsity of this cluster indicates that requirement-driven security analysis remains underdeveloped.

Synthesis of Cluster Roles

Taken together, the clusters reveal three partially aligned perspectives: system-centric security, project- and risk-oriented analysis, and requirements-based abstraction. While conceptual bridges exist, persistent fragmentation confirms the absence of cohesive frameworks integrating system-level attributes, requirement decomposition, and project-level risk evaluation.

6. Results

6.1. Trends Emerging from the Bibliometric Analysis

The bibliometric analysis shows rapid growth after 2019 followed by stabilization, indicating a shift from exploratory research to methodological consolidation. This is reflected in increasing reliance on secure SDLC guidance and standards-based frameworks, alongside a gradual rise in quantitative and model-driven risk assessment. Research remains geographically concentrated in countries with mature software and cybersecurity ecosystems, which shape dominant methodological perspectives.

The field is still led by computer science and software engineering, reinforcing a technical framing of security analysis, while organizational and socio-technical perspectives remain marginal. Thematic patterns indicate a gradual shift from system- and information-centered security towards project- and requirements-oriented approaches, although integration across these strands remains limited.

These findings align with earlier bibliometric studies reporting strong post-2018 growth, increasing dependence on secure SDLC frameworks, and persistent separation between technical security research and project-level decision processes (Khan et al., 2020; Williams et al., 2025). Compared with broader software-security reviews, this study also identifies a stronger emphasis on governance-oriented standards (ISO/NIST) and weaker integration of requirements engineering and quantitative risk modelling.

6.2. Literature Gaps with Direct Implications

Despite this consolidation, the literature shows persistent structural gaps with theoretical and practical implications. Cyber risk assessment and requirements engineering remain weakly integrated, so security requirements are often defined without explicit links to prioritized risk, limiting project-level usefulness. Security attributes from the CIA triad are widely referenced but rarely operationalized in analytical models supporting structured evaluation across decomposed requirements or development artifacts.

Many approaches also remain anchored at the organizational or governance level, providing strategic guidance but little operational support for software projects, especially in early

lifecycle phases where design trade-offs are critical. In addition, SME constraints are seldom treated as design assumptions, with frameworks often presuming expertise and tooling unrealistic for resource-constrained organizations.

Similar gaps are reported in previous bibliometric and systematic reviews, which highlight the dominance of technically focused approaches and limited integration of organizational constraints and early-phase requirements analysis (Khan et al., 2020; Leppänen et al., 2022). Their recurrence across independent studies suggests this fragmentation is structural rather than dataset specific.

6.3. How These Gaps Justify New Research

These gaps justify further research on project-level frameworks integrating cyber risk assessment with requirements engineering. Requirement-driven approaches combining systematic decomposition, CIA mapping, and structured risk scoring provide a basis for consistent prioritization and traceability across software artifacts (Fathabadi et al., 2024; Jiang et al., 2023).

This integration is particularly important for SMEs, where security measures must be feasible under resource constraints and growing regulatory pressure. Future frameworks should therefore prioritize lightweight design, tool support, and automation.

Recent studies also highlight the potential of emerging technologies, especially artificial intelligence for automated threat modelling, and risk prediction, and DevSecOps pipelines for continuous assessment and compliance verification (Bedoya et al., 2024). Together, these directions enable scalable, requirement-centered, and resource-efficient security analysis for modern software projects.

7. Conclusions

This study shows that research on security analysis in software development has grown steadily and increasingly converges around secure SDLC models and ISO/NIST standards, indicating methodological consolidation. However, risk assessment, requirements engineering, and security attributes remain weakly integrated, and most frameworks operate at the

governance level rather than at the project level. SMEs are particularly underrepresented despite their high-risk exposure.

Overall, the literature still lacks cohesive, project-level frameworks integrating requirement decomposition, security attributes, and structured risk assessment, highlighting the need for lightweight and operational approaches suited to real-world software development constraints.

References

Assal, H., & Chiasson, S. (2018). Security in the software development lifecycle. In *Fourteenth symposium on usable privacy and security (SOUPS 2018)* (pp. 281-296).

Bedoya, M., Palacios, S., Díaz-López, D., Laverde, E., & Nespoli, P. (2024). Enhancing DevSecOps practice with large language models and Security Chaos Engineering. *International Journal of Information Security*, *23*(6), 3765–3788. https://doi.org/10.1007/s10207-024-00909-w

Brown, C., & Thorpe, S. (2023). Towards Developing a formal model for tracking Cyber-Security Investments in Jamaica. In *SoutheastCon 2023*, 883–888.

El Amin, H., Samhat, A. E., Chamoun, M., Oueidat, L., & Feghali, A. (2024). An integrated approach to cyber risk management with cyber threat intelligence framework to secure critical infrastructure. *Journal of Cybersecurity and Privacy*, *4*(2), 357–381. https://doi.org/10.3390/jcp4020018

El-Hajj, M., & Mirza, Z. A. (2024). ProtectingSmall and Medium Enterprises: A specialized cybersecurity risk assessment framework and tool. *Electronics*, *13*(19), 3910. https://doi.org/10.3390/electronics13193910

Fathabadi, A. S., Snook, C., Dghaym, D., Hoang, T. S., Alotaibi, F., & Butler, M. (2025). Systematic hierarchical analysis of requirements for critical systems. *Innovations in Systems and Software Engineering*, *21*(2), 569–593. https://doi.org/10.1007/s11334-024-00551-8

Granata, D., & Rak, M. (2024). Systematic analysis of automated threat modelling techniques: Comparison of open-source tools. *Software Quality Journal*, *32*(1), 125–161. https://doi.org/10.1007/s11219-023-09634-4

Jiang, Y., Jeusfeld, M. A., Ding, J., & Sandahl, E. (2023). Model-based cybersecurity analysis: Extending enterprise modeling to critical infrastructure cybersecurity. *Business & Information Systems Engineering*, *65*(6), 643–676. https://doi.org/10.1007/s12599-023-00811-0

Khan, R., Khan, S., Ilyas, M., & Idris, M. (2020). *The State of the Art on Secure Software Engineering*. 24th International Conference on Evaluation and Assessment in Software Engineering, Pages 487 - 492. https://doi.org/10.1145/3383219.3383290

Leppänen, T., Honkaranta, A., & Costin, A. (2022). Trends for the DevOps security. A systematic literature review. Em *Lecture Notes in Business Information Processing* (pp. 200–217). Springer International Publishing. https://doi.org/10.1007/978-3-031-11510-3_12

McGraw, G. (2019). *Software security: Building security in* (2nd ed.). Addison-Wesley.

Quiñones, D., Rusu, C., & Rusu, V. (2018). A methodology to develop usability/user experience heuristics. *Computer Standards & Interfaces*, *59*, 109–129. https://doi.org/10.1016/j.csi.2018.03.002

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical Research Ed.)*, *372*, n71. https://doi.org/10.1136/bmj.n71

PRISMA statement. (no date). PRISMA Statement. Retrieved December 3, 2025, from https://www.prisma-statement.org/

Radanliev, P., De Roure, D., Nurse, J. R. C., et al. (2020). *Cyber risk in IoT systems*. IEEE Access, 8, 125249–125266. https://doi.org/10.20944/PREPRINTS201903.0104.V1

Rizvi, S., Zwerling, T., Thompson, B., Faiola, S., Campbell, S., Fisanick, S., & Hutnick, C. (2023). A modular framework for auditing IoT devices and networks. *Computers & Security*, *132*(103327), 103327. https://doi.org/10.1016/j.cose.2023.103327

Williams, L., Benedetti, G., Hamer, S., Paramitha, R., Rahman, I., Tamanna, M., Tystahl, G., Zahan, N., Morrison, P., Acar, Y., Cukier, M., Kästner, C., Kapravelos, A., Wermke, D., & Enck, W. (2025). *Research Directions in Software Supply Chain Security*. https://doi.org/10.1145/3714464

https://doi.org/10.58086/yv38-0307

ARTIFICIAL INTELLIGENCE–BASED SUPER NODES FOR REAL-TIME THREAT DETECTION IN DISTRIBUTED ENVIRONMENTS BIBLIOMETRIC ANALYSIS

José Lopes[1]✉, **Mário Dias Lousã**[1,2], **José Carlos Morais**[1,3]

[1] Instituto Superior Politécnico Gaya (ISPGAYA), Portugal.
[2] Insight - Piaget Research Center for Ecological Human Developement, Portugal.
[3] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal.
✉ Corresponding authors**:** ispg2021101231@ispgaya.pt

Abstract

The widespread adoption of distributed systems, driven by the growth of the Internet of Things (IoT), edge computing, and cloud infrastructure, has substantially expanded the attack surface of modern digital ecosystems. These environments, characterized by high heterogeneity, large data volumes, and stringent latency requirements, make real-time threat detection a complex task. Traditional, predominantly centralized security mechanisms reveal clear limitations in scalability and response time in the face of increasingly dynamic attack patterns. In this context, Artificial Intelligence (AI) and Machine Learning have emerged as essential enablers for more effective intrusion detection. At the same time, the concept of "super nodes" is gaining prominence: strategically positioned network elements with enhanced computational capabilities that act as intelligent intermediaries between edge devices and the central cloud. This study presents a bibliometric analysis of the use of AI-based super nodes for real-time threat detection. The analysis focuses on a sample of 300 publications indexed in the Lens.org database (2015–2025), selected according to the PRISMA 2020 guidelines. Through descriptive indicators and network analysis (such as keyword co-occurrence), research trends, thematic structures, and emerging directions in this field are identified.

Keywords: Artificial Intelligence; Machine Learning; IoT Security; Edge Intelligence; Bibliometric Analysis.

1. Introduction

Digital systems have undergone a profound transformation toward distributed architectures, supported by the proliferation of IoT devices, edge computing, and cloud. While these

technologies enable critical services, from healthcare to smart cities, they also introduce a substantial risk: a vastly expanded attack surface (Li et al., 2018; Conti et al., 2018).

The challenge lies in the nature of IoT and edge environments. They are heterogeneous, resource-constrained, and demand ultra-low latency. Traditional, centralized security models struggle here; they simply cannot scale or adapt fast enough to stop real-time threats (Roman et al., 2018; Zhou et al., 2019). The consequences are evident in the rise of devastating DDoS botnets and ransomware attacks targeting industrial systems (Shafiq et al., 2024; Singh et al., 2021; Vakulov, 2025).

In response, Artificial Intelligence (AI) and Machine Learning have emerged as essential tools. Moving beyond static rules, these models offer the ability to detect anomalies dynamically in large data streams (Al-Garadi et al., 2020; Goranin et al., 2024). However, software needs the right architecture. This is where 'super nodes' become relevant. Acting as powerful, intelligent intermediaries between the edge and the cloud, super nodes run AI algorithms locally to detect threats with minimal latency (Nguyen & Reddi, 2021; Zhou et al., 2019).

Yet, despite the clear synergy between AI and super nodes, scientific literature on the subject is scattered across engineering, security, and data science fields. This lack of cohesion makes it difficult to compare solutions or grasp the full scope of innovation (Goranin et al., 2024). To bridge this gap, this article provides the first comprehensive bibliometric analysis of AI-based super nodes for real-time threat detection, structuring existing knowledge and charting the course for future research.

2. Methodology

This research adopts a bibliometric approach with the aim of systematically and structurally analyzing the scientific production related to the use of AI-based super nodes for real-time threat detection in distributed environments. The choice of bibliometric methodology is justified by the need to synthesize a broad, interdisciplinary, and rapidly growing body of literature. This allows the identification of research trends, scientific collaboration structures, dominant thematic areas, and conceptual gaps that would be difficult to observe through traditional narrative reviews (Van Eck & Waltman, 2010).

The methodological design followed the PRISMA 2020 guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), widely recognized as a benchmark of

good practices for systematic and bibliometric reviews (Page et al., 2021; PeerJ Computer Science, 2024). The application of these guidelines allowed for the transparent and reproducible structuring of all phases of the study selection process, from the initial identification of records to the definition of the final sample. The complete process of identifying, screening, eligibility, and inclusion of studies is summarized in the flowchart presented in Figure 1.

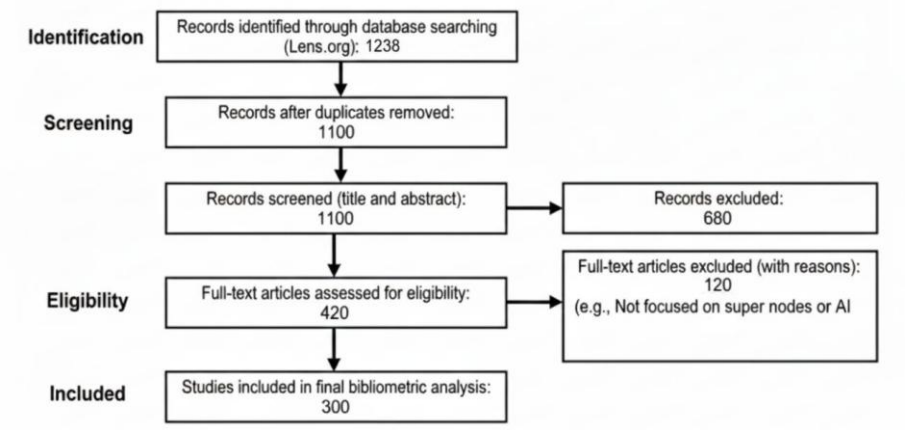

Fig 1. PRISMA 2020 flow diagram
Source: Authors' elaboration.

2.1. Data Source

The Lens.org database was selected as the primary source for collecting scientific literature. This choice is based on its broad interdisciplinary coverage, integrating relevant publications in the areas of cybersecurity, computer science, engineering, telecommunications, and distributed systems. Furthermore, Lens provides rich and structured bibliographic metadata, allowing the direct export of information compatible with bibliometric analysis tools, facilitating the replication of the study. When compared to commercial databases such as Scopus or Web of Science, the Lens platform presents specific advantages for the context of this work, namely open access and strong representation of technical literature associated with the Internet of Things and edge computing. However, it is recognized that the exclusive use of a single database may introduce limitations in terms of coverage of certain publications, an aspect that is critically considered in section four (Discussion).

2.2. Search Strategy

The search strategy was designed to capture studies that combined three central dimensions of the domain under analysis: (i) the application of Artificial Intelligence or Machine Learning, (ii) the focus on threat or intrusion detection, and (iii) the use in distributed environments, including IoT, edge computing, and hierarchical network architectures.

For this purpose, a Boolean search structure was applied to titles, abstracts, and keywords, according to the following general scheme:

("artificial intelligence" OR "machine learning") AND

("threat detection" OR "intrusion detection") AND

("IoT" OR "edge computing" OR "distributed systems")

The period considered encompassed publications between 2015 and 2025, reflecting the significant growth of research in IoT and edge intelligence from the second half of the 2010s, as well as the maturation of deep learning techniques applied to cybersecurity. No geographical or publication type restrictions were imposed, provided the studies were relevant to the topic under analysis.

2.3. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were defined beforehand, with the aim of ensuring thematic coherence and methodological quality of the final sample. Studies applying Artificial Intelligence or machine learning techniques to the detection of threats, intrusions, or anomalies in distributed environments were included, as well as works describing architectures involving nodes with enhanced analytical capabilities, explicitly or implicitly associated with the concept of super nodes.

On the other hand, studies unrelated to cybersecurity, those exclusively addressing centralized architectures without a distributed component, or those presenting insufficient metadata for a consistent bibliometric analysis were excluded. The clear definition of these criteria contributed to reducing biases in the selection process and ensuring the scientific relevance of the sample analyzed.

2.4. Study Selection Process

The selection process followed the PRISMA workflow. The initial search resulted in the identification of 1,238 records. After removing duplicates, 1,100 publications remained for title and abstract analysis. At this stage, 680 studies were excluded due to lack of thematic relevance.

The remaining 420 articles were analyzed in full text, leading to the additional exclusion of 120 publications that did not fully meet the defined inclusion criteria, namely because they did not address AI techniques or distributed architectures compatible with the concept of super nodes. The final sample comprised 300 studies, forming the basis of the bibliometric analysis presented in this work, as illustrated in Figure 1.

2.5. Bibliometric Analysis Techniques and Tools

The bibliometric analysis was conducted in two complementary phases. In the first phase, descriptive indicators were extracted using the analytical tools provided by Lens.org, including the temporal evolution of publications, distribution by scientific sources, disciplinary areas, and countries of affiliation of the authors.

In the second phase, the VOSviewer software was used to construct bibliometric networks, namely keyword co-occurrence networks and co-authorship networks. For the generation of these networks, a minimum threshold of five occurrences per term or author and a minimum link strength of two were defined, ensuring the analytical relevance and readability of the visualizations (Van Eck & Waltman, 2010). The interpretation of the identified clusters was carried out in conjunction with the concept of super nodes, allowing the observed bibliometric trends to be related to the conceptual and technological evolution of the domain under study.

3. Bibliometric Results

This section presents the main results of the bibliometric analysis performed on the final sample of 300 publications, selected according to the process described in the Methodology section. The presentation of the results follows the same logic and numbering of the figures

in the original article, ensuring structural coherence and facilitating the reading and validation of the data. The results are organized into different analytical dimensions, including thematic analysis of the literature, patterns of scientific collaboration, publication sources, geographical distribution of research, and temporal evolution associated with scientific impact.

3.1. Keyword Co-occurrence and Thematic Clusters

Keyword co-occurrence analysis allowed us to identify the main research themes and understand the conceptual structure of the domain under study. The keyword network, shown in Figure 2, highlights the existence of several thematic clusters, reflecting different lines of research related to threat detection based on Artificial Intelligence in distributed environments.

The dominant cluster is centered on terms such as “machine learning”, “deep learning”, “intrusion detection”, and “anomaly detection”, representing the technical core of the domain. These concepts reflect the strong reliance on machine learning techniques for detecting malicious behavior. A second cluster aggregates terms related to IoT, edge computing, distributed systems, and fog computing, highlighting the growing concern about security in distributed environments with limited computing resources. The interconnection between these clusters suggests the emergence of hybrid architectures, in which super nodes assume an intermediary role of aggregation and intelligent analysis between edge devices and cloud infrastructures.

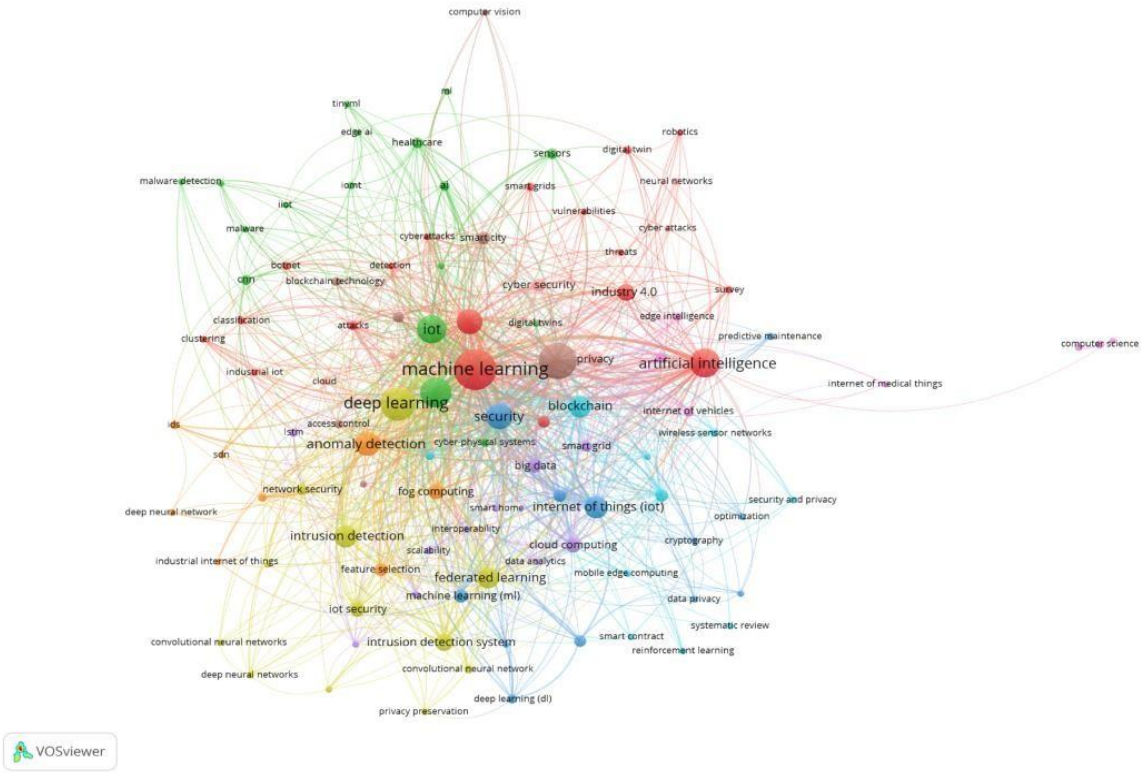

Fig 2. Keyword co-occurrence network generated, highlighting the main thematic clusters in the analyzed literature
Source: VOSviewer (2025).

3.2. Author Collaboration and Co-authorship Networks

The analysis of co-authorship networks allows us to understand the patterns of collaboration between researchers and institutions in the analyzed domain. Figure 3 presents the co-authorship network, highlighting the existence of several relatively isolated research groups with limited interconnection between them.

Despite the presence of some more central authors and institutions, the network is characterized by significant fragmentation, suggesting that the field is still in a phase of scientific consolidation. This fragmentation could be explained by the interdisciplinary nature of the topic, which crosses areas such as cybersecurity, networks, and Artificial Intelligence, as well as by the absence of a widely accepted formal conceptualization of the concept of super nodes.

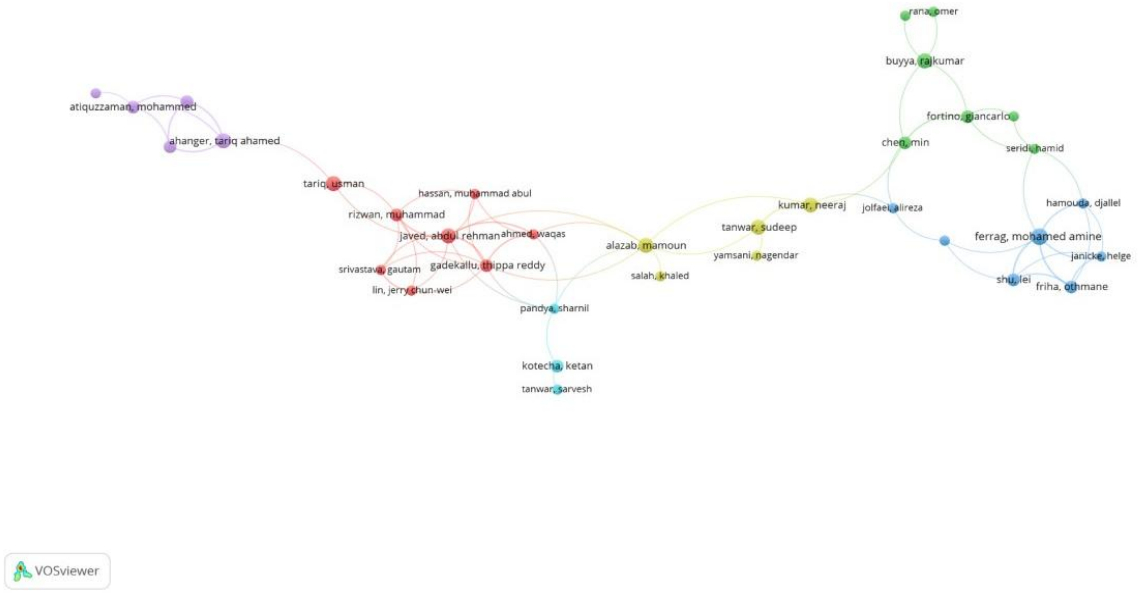

Fig 3. Co-authorship network illustrating collaboration clusters among authors publishing on AI-based threat detection in distributed environments
Source: VOSviewer (2025).

3.3. Publication Venues and Document Types

Analysis of publication sources reveals that scientific output is distributed across a diverse set of journals and conferences, primarily in the fields of computer science, engineering, and telecommunications. Figure 4 shows the distribution of publications among the main scientific journals, highlighting the predominance of specialized technical sources.

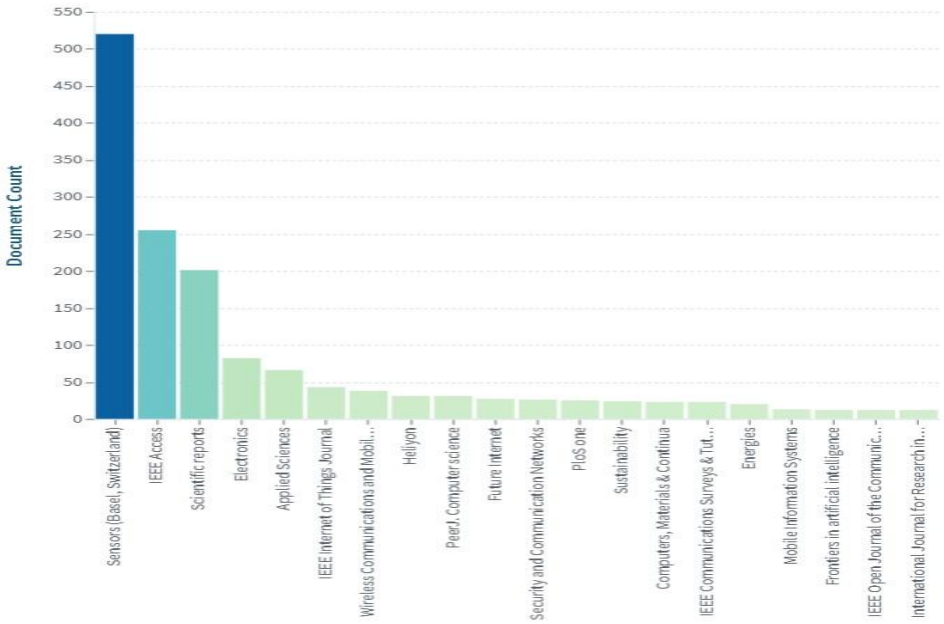

Fig 4. Distribution of publications by major journals
Source: The Lens (2025).

Additionally, Figure 5 illustrates the distribution of document types included in the sample, highlighting the significant weight of conference papers and journal articles, reflecting the dynamic and emerging nature of the field under study.

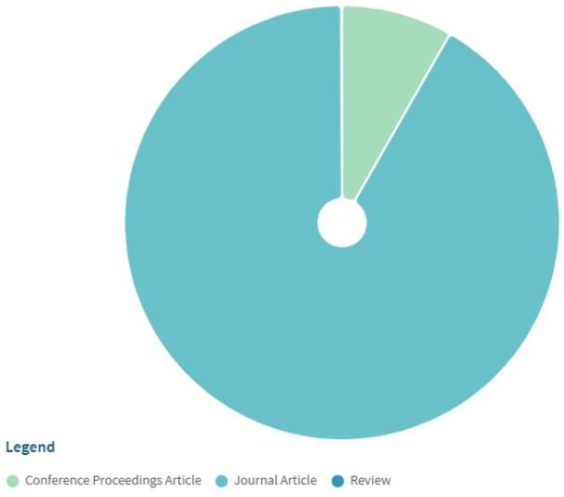

Fig 5. Distribution of document types (journal articles, conference papers, and reviews)
Source: The Lens (2025).

3.4. Geographical Distribution of Research

Analysis of the geographical distribution of publications indicates that research in this field is led by a relatively small number of countries. As illustrated in Figure 6, scientific output is concentrated mainly in China, the United States, India, and several European countries.

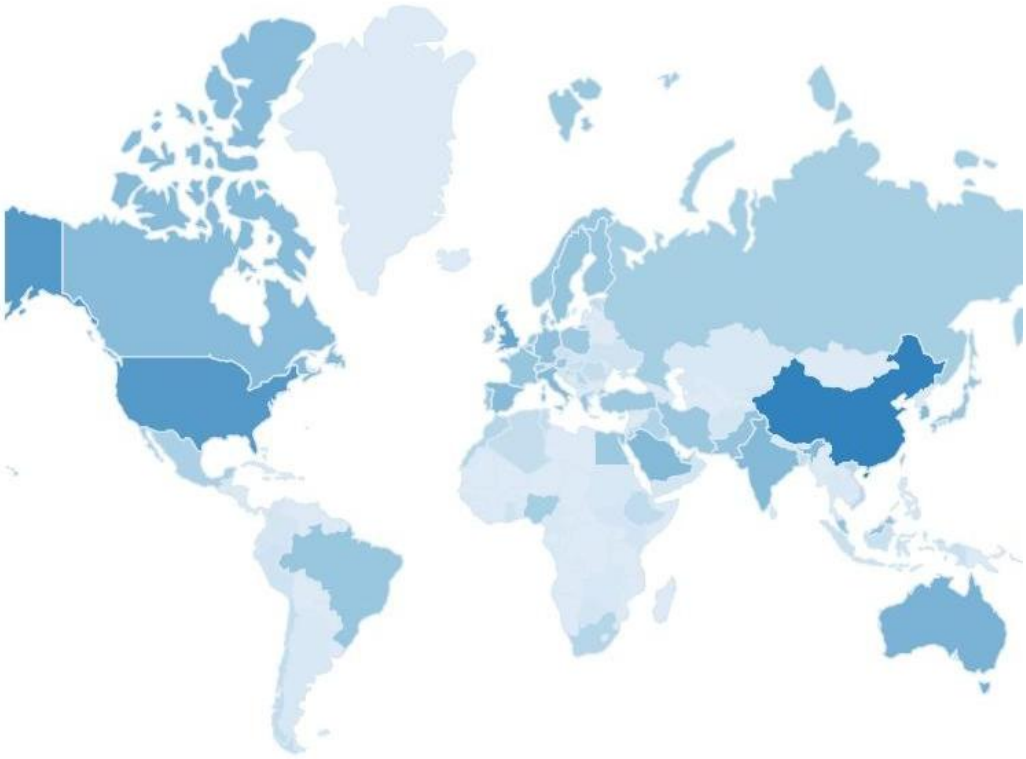

Fig 6. Global geographical distribution of publications by country
Source: The Lens (2025).

This concentration reflects these countries' significant investment in emerging technologies, IoT infrastructure, and artificial intelligence research. However, the analysis also highlights limited international collaboration, reinforcing the perception of fragmentation in the field and pointing to future opportunities for transnational collaborative research.

3.5. Temporal Evolution and Citation Impact

The temporal evolution of scientific production, associated with the impact of publications, is represented in Figure 7. This figure combines the temporal dimension with impact metrics,

allowing us to observe not only the growth in the number of publications over time but also their scientific relevance measured through citations.

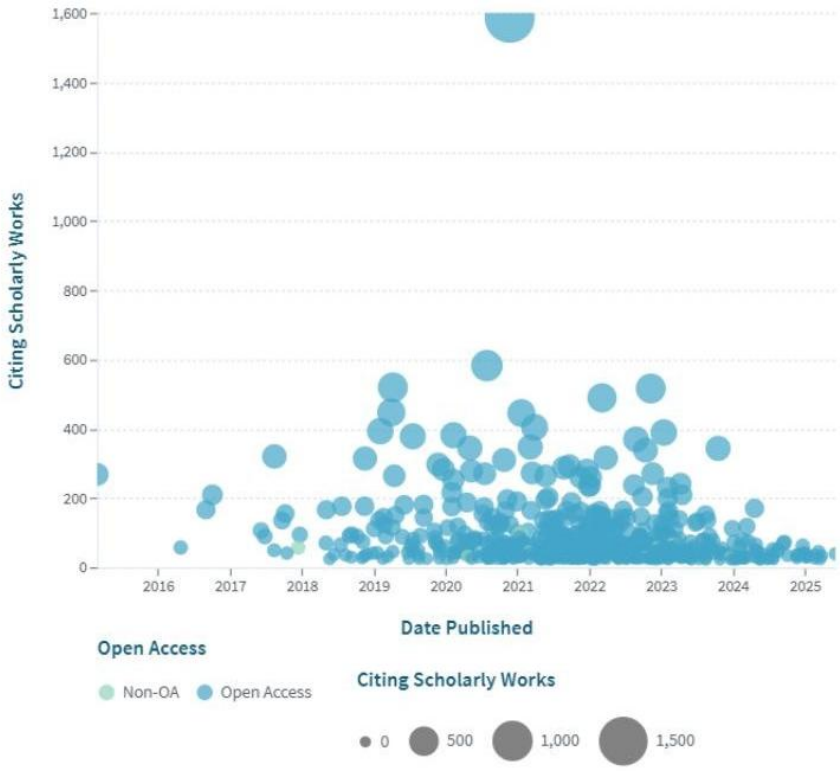

Fig 7. Temporal evolution of publications and citation impact over time
Source: The Lens (2025).

The results show a significant growth in research since 2018, coinciding with the maturation of IoT and edge computing technologies. The increased scientific impact associated with the most recent publications suggests a progressive consolidation of the domain and a growing interest in intelligent distributed architectures, in which super nodes play a central role.

3.6. Scientific Domains Involved

Finally, the analysis of the scientific domains of the included publications reveals a strong concentration in the areas of computer science, engineering, and information technologies. Figure 8 presents a visualization of the main scientific domains associated with the analyzed literature.

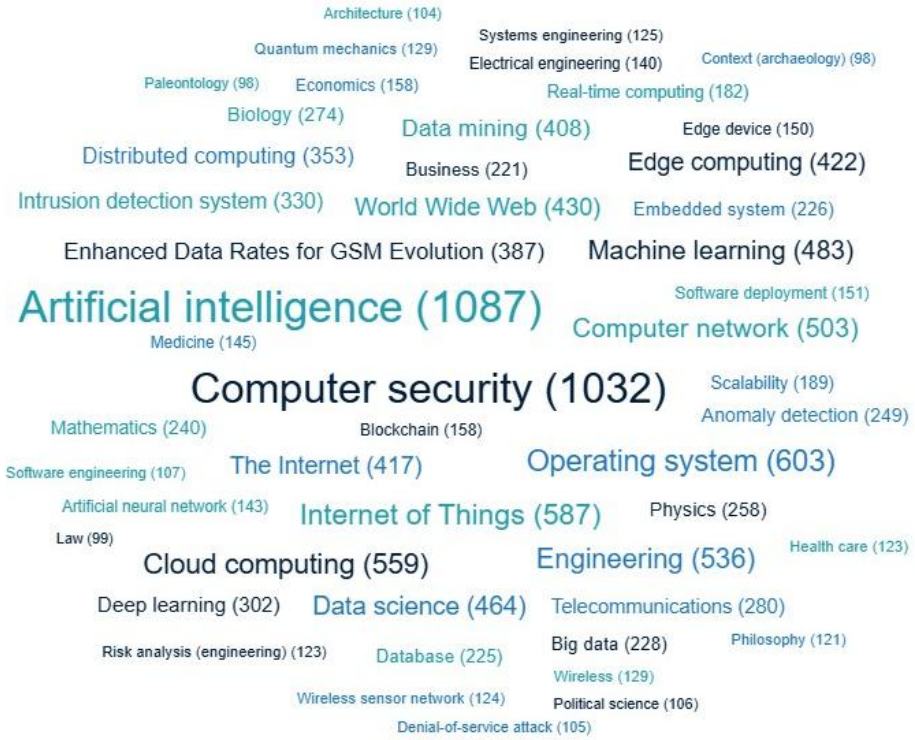

Fig 8. Scientific domains associated with the analyzed publications represented as a word cloud or domain-frequency visualization
Source: The Lens (2025).

This technical focus indicates that research on super nodes and threat detection is still in a phase of technological consolidation, and there is room for future, more integrated approaches that consider organizational, regulatory, and social dimensions.

4. Discussion

The bibliometric analysis carried out allows us to draw a set of relevant conclusions about the scientific evolution of the field of threat detection based on Artificial Intelligence in distributed environments, as well as about the emerging role of super nodes in this context. The results show a rapidly growing but still fragmented research field, both at the conceptual level and at the level of scientific collaboration, confirming several of the gaps identified in the literature.

First, the temporal evolution and impact of publications indicate that scientific interest in this field has intensified significantly since 2018. This growth coincides with the maturation of Internet of Things and edge computing technologies, as well as advances in deep learning techniques applied to cybersecurity. This technological convergence has created the

necessary conditions for the development of more distributed and intelligent architectures, capable of responding to the scalability and latency limitations of traditional centralized models. The increasing relevance of the most recent publications, reflected in their scientific impact, suggests that the topic is in a phase of consolidation and recognition within the academic community.

Keyword co-occurrence analysis revealed a relatively well-defined thematic structure, organized into clusters that reflect the main research lines of the domain. On the one hand, there is a technical core strongly focused on machine learning and deep learning techniques for intrusion and anomaly detection. On the other hand, a set of themes associated with distributed environments emerges, such as IoT, edge computing, and hybrid architectures. The articulation between these clusters suggests that research has been evolving towards the integration of advanced analytical capabilities and distributed architectures, creating a clear conceptual space for the emergence of super nodes as intermediate elements of intelligence.

However, despite this thematic convergence, the literature reveals an absence of a formal and widely accepted conceptualization of the concept of super nodes. Many studies describe architectures that incorporate nodes with enhanced computational capabilities and local aggregation and analysis functions, but they do so use diverse and sometimes inconsistent terminology. This conceptual fragmentation hinders systematic comparison between approaches, as well as the construction of reference models that can guide future investigations and practical implementations.

The identified patterns of scientific collaboration reinforce this perception of fragmentation. The analysis of co-authorship networks reveals the existence of multiple relatively isolated research groups, with reduced interconnection between them. This configuration can be partially explained by the interdisciplinary nature of the domain, which crosses areas such as cybersecurity, networks, artificial intelligence, and systems engineering. However, the absence of strongly interconnected scientific communities may also limit the consolidation of common approaches and the evolution of the field towards more mature and standardized solutions.

The geographical distribution of scientific production also reveals a significant concentration of research in countries with strong investment in emerging technologies, such as China, the United States, and several European countries. Although this concentration is expected, considering the resources and infrastructure available in these contexts, the analysis also shows limited international collaboration. This aspect represents both a current limitation

and a future opportunity, namely for the development of collaborative projects that allow the validation of super node-based architectures in different geographical and operational contexts.

Regarding the predominant scientific domains, the results indicate that research is strongly anchored in technical areas, such as computer science and engineering. The limited presence of approaches originating from areas such as management, public policy, or regulatory frameworks suggests that the domain is still in a phase of technological consolidation. As super node-based solutions evolve into real-world application contexts, it will be crucial to integrate these dimensions, especially in critical environments where privacy, accountability, and regulatory compliance issues play a central role.

In summary, the results of the bibliometric analysis confirm that super nodes constitute an emerging concept with high potential to address the challenges of threat detection in distributed environments. However, the consolidation of this paradigm requires further effort in conceptual systematization, empirical validation, and scientific collaboration. This research contributes to this effort by structurally mapping the state of the art, identifying relevant gaps, and providing a solid foundation for future research in this field.

5. Conclusion

This article aimed to systematically analyze the scientific production related to the use of AI-based super nodes for real-time threat detection in distributed environments. To this end, a bibliometric analysis of 300 scientific publications was conducted, selected according to the PRISMA 2020 guidelines and extracted from the Lens.org database. This allowed for mapping the evolution of the domain, identifying research trends, thematic structures, and relevant gaps in the literature.

The results show a significant growth in research in this field since 2018, reflecting the convergence between the maturation of Internet of Things technologies, edge computing, and advances in machine learning and deep learning techniques applied to cybersecurity. The thematic analysis revealed well-defined clusters, centered, on the one hand, on intrusion detection techniques based on machine learning and, on the other hand, on distributed and hybrid architectures. This thematic convergence reinforces the growing relevance of

architectures that integrate intelligent intermediate nodes, conceptualizable as super nodes, capable of mitigating limitations associated with centralized security models.

Despite this progressive growth and maturation, bibliometric analysis has also revealed a set of structural limitations in the field. In particular, the literature appears fragmented, both at the conceptual level and at the level of scientific collaboration, with an absence of a formal and widely accepted definition of the concept of super nodes. Many studies address functionally similar architectures but resort to diverse and sometimes inconsistent terminology, hindering the systematic comparison of approaches and the consolidation of scientific knowledge.

Additionally, the patterns of scientific collaboration and the geographical distribution of publications suggest that research is concentrated in certain countries and research groups, with relatively small levels of international collaboration. This fragmentation could limit the empirical validation of proposed solutions in diverse operational contexts, as well as delay the emergence of widely recognized reference models or best practices.

In this context, super nodes emerge as a promising paradigm for threat detection in distributed environments, offering an intermediate approach between resource-limited edge devices and centralized cloud platforms. Their ability to aggregate data, perform local AI-based inference, and make decisions in reduced time makes them particularly suitable for scenarios that demand rapid responses, scalability, and resilience.

Future research areas include the need to develop a formal and standardized conceptualization of the super node concept, as well as conducting empirical studies to evaluate their performance in real or simulated environments. Future investigations could also explore the integration of super nodes with distributed orchestration mechanisms, federated learning techniques, and adaptive security approaches. Finally, it will be relevant to incorporate organizational, regulatory, and ethical perspectives to ensure that the adoption of these architectures occurs in a secure, transparent manner, aligned with legal and social requirements.

In summary, this work contributes to the systematization of the state of the art on super nodes and threat detection based on Artificial Intelligence, providing a solid foundation for advancing research in this emerging and technologically relevant field.

References

Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for Internet of Things (IoT) security. *IEEE Communications Surveys & Tutorials, 22*(3), 1646–1685. https://doi.org/10.1109/COMST.2020.2988293

Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities. *Future Generation Computer Systems*, *78*, 544-546.

Goranin, N., Hora, S. K., & Čenys, H. A. (2024). A bibliometric review of intrusion detection research in IoT: Trends, collaboration, and thematic evolution. *Electronics, 13*(16), 3210. https://doi.org/10.3390/electronics13163210

Li, C., Xue, Y., Wang, J., Zhang, W., & Li, T. (2018). Edge-oriented computing paradigms: A survey on architecture design and system management. *ACM Computing Surveys (CSUR)*, *51*(2), 1-34. https://doi.org/10.1145/3154815

Nguyen, T. T., & Reddi, V. J. (2023). Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(8), 3779–3795. https://doi.org/10.1109/TNNLS.2021.3121870

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Roman, R., Lopez, J., & Mambo, M. (2018). Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, *78*, 680-698.

Shafiq, M., Tian, Z., Liu, Y., Aljuhani, A., & Li, Y. (2024). ESC&RAO: Enabling seamless connectivity resource allocation in tactile IoT for consumer electronics. *IEEE Transactions on Consumer Electronics*, *70*(3), 5506–5515. https://doi.org/10.1109/tce.2023.3327136

Singh, P., Kaur, A., & Singh, D. (2021). Edge-centric network intrusion detection using deep learning. *arXiv Preprint*, arXiv:2102.01873. https://arxiv.org/abs/2102.01873

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE, 107*(8), 1738–1762. https://doi.org/10.1109/JPROC.2019.2918951

https://doi.org/10.58086/w1yd-0829

GENERATIVE AI MUTABILITY IN CYBERSECURITY: A BIBLIOMETRIC REVIEW

Pedro Oliveira[1]✉, **Mário Dias Lousã**[1,2], **José Carlos Morais**[1,3]

[1] Instituto Superior Politécnico Gaya (ISPGAYA), Portugal.
[2] Insight - Piaget Research Center for Ecological Human Developement, Portugal.
[3] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal.
✉ Corresponding authors**:** ispg2024103478@ispgaya.pt

Abstract

The expansion of generative AI (GenAI) is forcing us to rethink cybersecurity, expanding both defensive automation and scalable offensive techniques. This bibliometric review maps the change driven by GenAI in cybersecurity through a PRISMA-guided selection of 154 documents from The Lens (20 December 2025). The current state is summarized by scientific mapping results (co-authorship, co-word, and co-citation networks, and thematic evolution) to identify dominant architectures, thematic clusters, and collaboration patterns, and implications for governance and auditing. We note the exponential growth of publications in 2022. We notice the trend. The authors group publications into several architectures: large language models (LLMs), generative networks (GANs), and diffusion models. These focus on common topics, (i) large-scale phishing and social engineering, (ii) mutability, obfuscation, and adversarial evasion of malware, and (iii) intrusion detection and cyber threat intelligence using synthetic data. Co-citation networks and keywords show that adversarial robustness, red teaming, and benchmarking are interconnected. We find that explainability and human-in-the-loop defense exist as minor but growing topics. One risk is the BlackMamba case, which transmits an LLM-assisted pipeline capable of generating more than 10,000 semantically identical but structurally distinct mutations per hour and achieving a 98.2% evasion rate against commercial EDR solutions. Risk mitigation should prioritize benchmarking and standardized reporting, continuous red teaming, and telemetry monitoring, incorporated into dynamic audit frameworks, supported by explicit international governance for high-risk GenAI cybersecurity applications.

Keywords: Adversarial AI; AI-driven cyber threats; Polymorphic and metamorphic malware; Synthetic data; Intrusion detection systems.

1. Introduction

Cybersecurity faces a rapidly evolving threat landscape in which adversaries iterate faster than defenders. GenAI [including GANs, large language models (LLMs), and diffusion models] alters this balance by reducing the cost of content generation, code transformation, and scalable automation (Gupta et al., 2023), thereby intensifying routine and advanced cyber operations. A concrete illustration of the mutability enabled by GenAI is BlackMamba, in which an LLM is used to continuously generate and rewrite a keylogger, helping it evade signature- and behavior-based controls such as endpoint detection and response (EDR) (Ibrar et al., 2025). Such cases exemplify how GenAI supports polymorphic and metamorphic behaviors that undermine traditional detection pipelines (Javaheri et al., 2021). This bibliometric review maps the research landscape on GenAI and mutability in cybersecurity, using n=154 documents (2018–20 December 2025) indexed in The Lens and selected under PRISMA. The objective is to quantify growth, identify influential sources and actors, and characterize dominant themes in defensive and offensive research trajectories. This bibliometric analysis investigates the following research questions:

RQ1: Identify scientific production over time, main sources, authors, and countries leading the field?

RQ2: Which thematic clusters and GenAI architectures dominate GenAI research in cybersecurity?

RQ3: How is the literature distributed between defensive applications and offensive/evasion-oriented research?

RQ4: What limitations and research gaps emerge for future work on GenAI mutability in cybersecurity?

This article is structured as follows. Section two summarizes the conceptual background on GenAI architectures and mutability. Section three details the PRISMA-guided bibliometric methodology, including the search strategy, inclusion criteria, and analytical tools. Section four presents the bibliometric results (production trends, sources, actors, and thematic structures). Section five discusses the implications, including attack-defense asymmetry and governance challenges. Section six concludes with limitations and future directions.

2. Literature review

GenAI is characterized by scalable content and code generation, enabling both defensive analytics and offensive capability. In cybersecurity, the same generative mechanisms that support detection and simulation can also be repurposed to evade controls through continuous variation (mutability) (Gazzan et al., 2025).

GenAI architectures appear in the corpus primarily as (i) GANs, which learn to synthesize realistic samples via generator–discriminator competition and are frequently used for synthetic security data and adversarial example generation (Mahmoudi et al., 2025; Pei et al., n.d.); (ii) LLMs and Transformer variants, which support Cyber Threat Intelligence (CTI), phishing and social-engineering automation, and code transformation/rewriting (Balasubramanian et al., 2025); and (iii) diffusion models, an emerging class increasingly used for higher-fidelity synthetic generation and augmentation (Yazdani et al., 2025). Across architectures, the security relevance is operational: models enable automation, scalability, and rapid iteration under constrained defender visibility. GANs: Composed of a generator and a discriminator in competition, GANs are widely used to generate adversarial attack traffic that "fools" the detector, forcing it to learn more robust patterns. In Internet of Vehicles (IoV) environments (Dunmore et al., 2023), GANs are crucial for simulating attacks and strengthening intrusion detection. GenAI facilitates the creation of malware that alters its signature to evade detection (Labaca-Castro, 2023).

Metamorphism via LLMs: "Uncensored" or jailbroken LLMs can rewrite the syntax of malicious code (e.g., renaming variables, inserting dead code) while maintaining the original semantics. This automates the creation of zero-day variants (Acosta-Bermejo et al., 2025).

Mutability refers to the capability of malicious artifacts to change form while preserving function. Polymorphism typically modifies surface-level representations (e.g., encryption/packing and variable changes) to break signatures, whereas metamorphism rewrites code structure to alter control flow and semantics (Guo, 2023). LLMs can accelerate metamorphic behavior by rewriting payloads on demand, including renaming, refactoring, and generating variant implementations. BlackMamba exemplifies LLM-enabled metamorphism: a keylogger is generated and iteratively modified to reduce detection likelihood, challenging EDR pipelines that rely on known behavioral patterns (Ibrar et al., 2025). In parallel, GAN-based approaches support adversarial attacks by producing near-indistinguishable perturbations that degrade malware or intrusion classifiers, exposing vulnerabilities in AI-based

defense (Liu et al., 2025). Defensively, GenAI is used to address data scarcity and imbalance through synthetic data generation for IDS training, attack simulation, and IoT/edge scenarios where labeled telemetry is limited (Kumar et al., 2025). However, dual-use nature implies that improved generation and simulation capabilities must be paired with rigorous evaluation, adversarial testing, and governance to prevent the same tooling from scaling offensive operations (Radanliev, 2025).

Adversarial attacks: GANs are used to create adversarial examples — samples of malware with minimal perturbations, invisible to humans but causing misclassification by deep learning models (Gaber et al., 2024).

3. Methodology

The final dataset consists of n=154 records after applying the PRISMA protocol (Figure 1). The data served as the basis for obtaining descriptive indicators, co-occurrence structures, and co-citation networks using Bibliometrix (R) and VOSviewer.

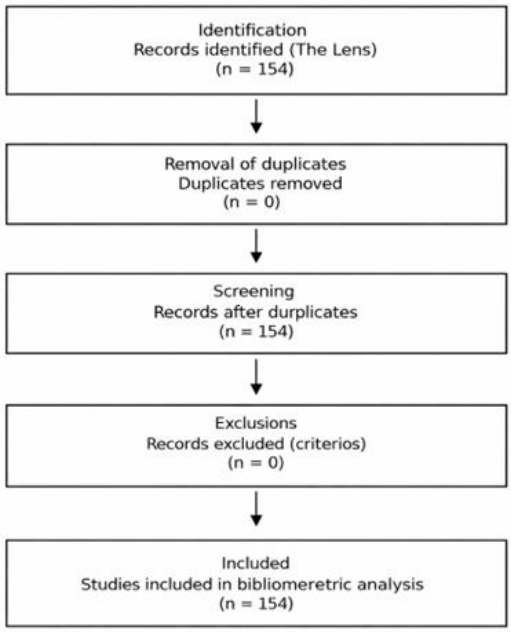

Fig 1. PRISMA diagram
Source: Authors' elaboration.

3.1. Data source and analysis period

The search was executed in The Lens on 20 December 2025. Data collection was carried out on The Lens platform, which aggregates scientific publications from multiple publishers and indexing databases (IEEE Xplore, ACM Digital Library, Scopus, SpringerLink), and the records were exported on the same date, yielding N=154 documents published between 2018 and 20 December 2025.

3.2. Research strategy

The search strategy was executed in the "Full Text" field of The Lens platform, using the following query: ("generative AI" AND "cybersecurity") OR ("generative AI" AND "generative adversarial networks") OR ("metamorphic" AND "generative AI") OR ("generative AI" AND "polymorphic") OR ("generative AI" AND "red team") OR ("generative AI" and "malware") OR ("generative adversarial networks" AND "metamorphic") OR ("generative adversarial networks" AND "polymorphic")]. The following filters were applied: "Document Type: Journal Article or Conference Article or Book Chapter"; "Language: English"; and "Period: 2018-20 December 2025". The subject matter: "Computer Science Applications, Artificial Intelligence, Information Systems, Computer Networks and Communications, General Computer Science, Molecular Biology, Software, Library and Information Sciences, General Engineering, Physical and Theoretical Chemistry". The exported records included complete bibliographic metadata, citations, author keywords, and abstracts.

3.3. Inclusion and exclusion criteria

We included studies published from early 2018 to 20 December 2025 (based on The Lens database) that explicitly address the use of generative models in creating threats or defense mechanisms, with experimental validation. Articles on generative AI unrelated to security were excluded, as were cybersecurity studies that only use classical discriminative techniques. Technical reports and non-peer-reviewed documents were also excluded, unless explicitly referenced in systematic reviews. No records were excluded during screening (n=0)

because the search string and filters were highly specific, yielding a corpus already aligned with the intersection of GenAI and cybersecurity.

3.4. Processing and tools

The files were cleaned and standardized in Microsoft Excel, removing possible duplicates and unifying variants of keywords and author names. The data was then analyzed in the Bibliometrix (R) package to generate indicators of scientific output, citations, sources, authors, countries, and term co-occurrence networks. VOSviewer was used to visualize network maps, density maps, and temporal overlap, while additional graphs were generated in The Lens itself.

4. Bibliometric Results

Analysis of empirical data reveals clear trends that answer research questions.

4.1. Scientific output and life cycle of the topic

Across 2018–20 December 2025, annual output accelerates markedly after 2022 (Figure 2), consistent with the diffusion of LLMs (e.g., ChatGPT) into mainstream security practice and research.

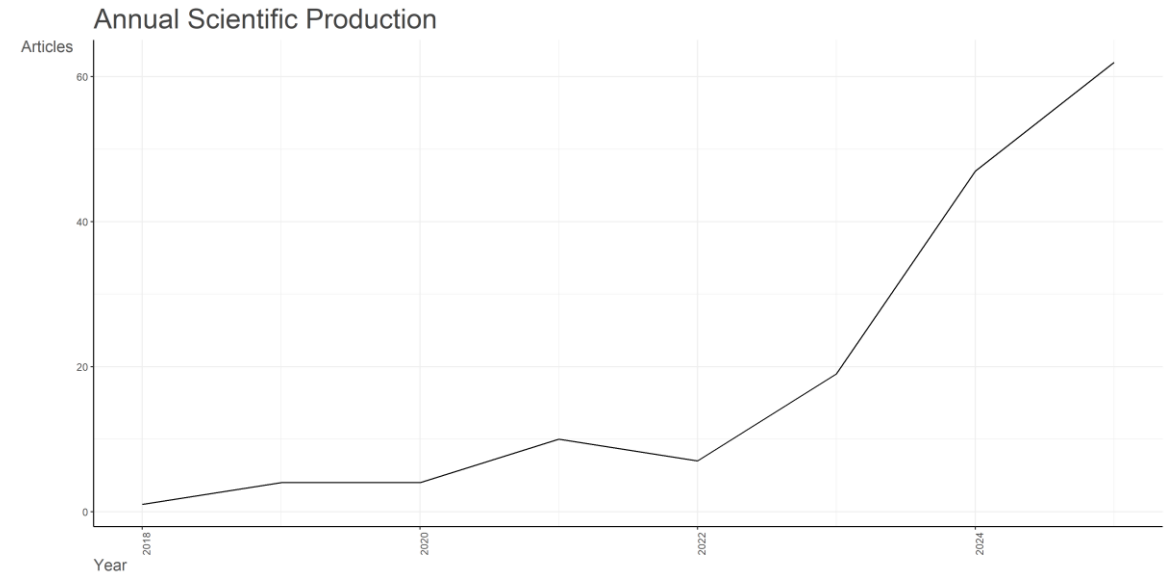

Fig 2. Annual scientific output
Source: output from Bibliometrix (2025).

The life-cycle indicators suggest a growth phase rather than maturation, while citation indicators (Figure 3) should be interpreted cautiously for recent years due to shorter accumulation windows. The decline in average citations per article in recent years reflects the temporal citation accumulation effect, not reduced topic relevance.

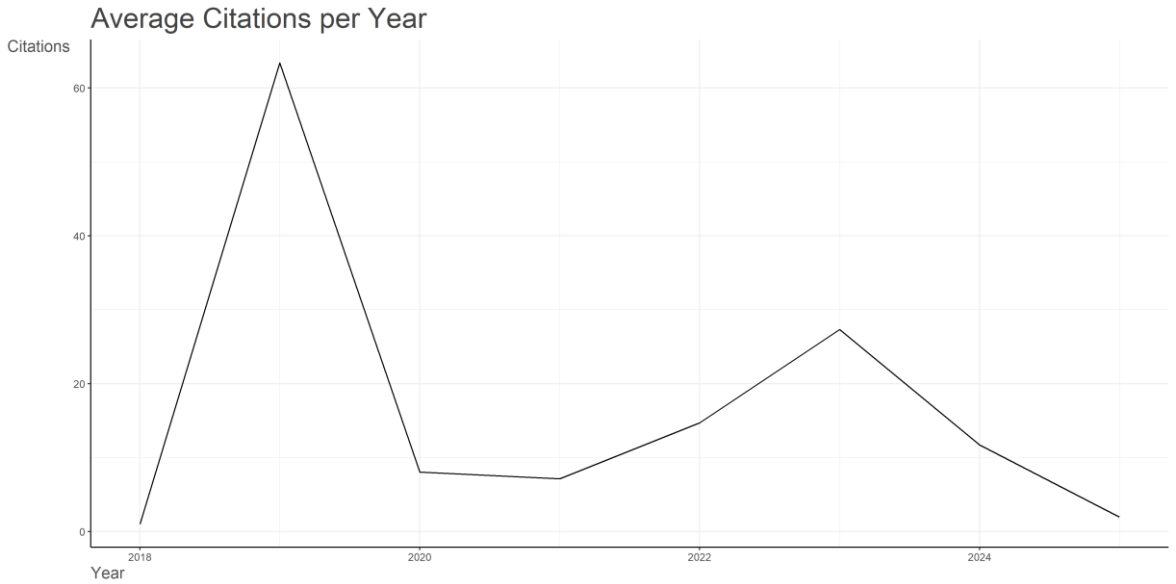

Fig 3. Graph showing average citations per article and per year
Source: output from Bibliometrix (2025).

4.2. Types of documents

The corpus is dominated by journal articles and conference proceedings, indicating both rapid dissemination and consolidation of findings (Figure 4).

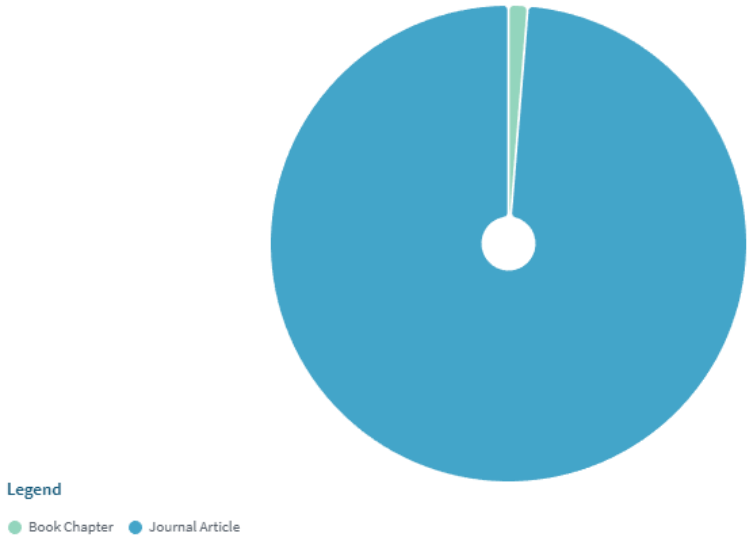

Fig 4. Distribution of publication types in the sample analyzed
Source: The Lens (2025).

4.3. Most active sources, and authors

Scientific production is concentrated in a small set of outlets and actors, representing a large share of publications, according to Bradford's Law (Figure 5).

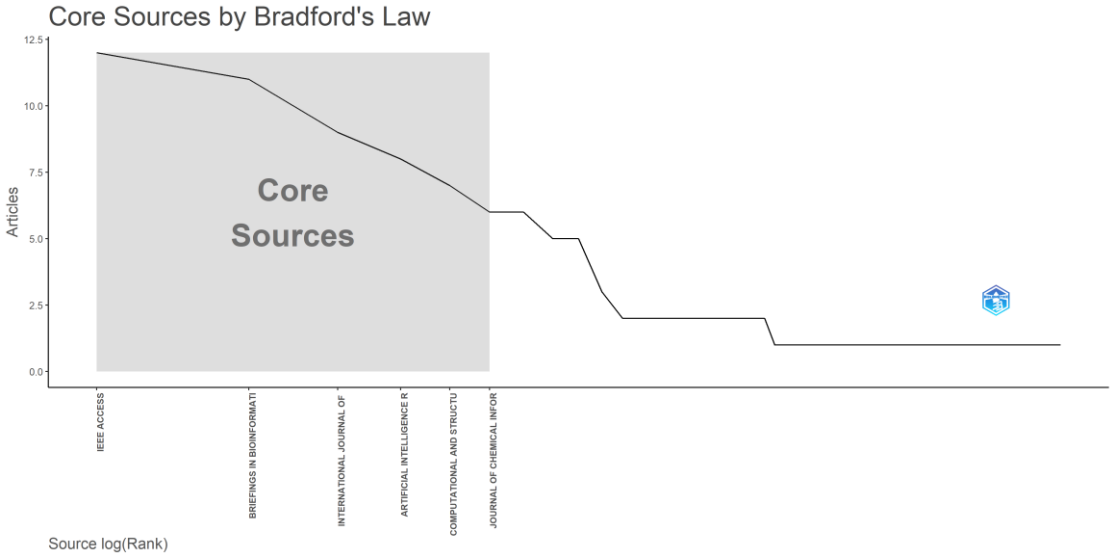

Fig 5. Core sources according to Bradford's Law

Source: output from Bibliometrix (2025).

IEEE Access emerges as the most influential journal, recording the highest metrics with an h-index of 8, a g-index of 12, and a total citation count of 1,333 (Table 1). These figures reflect a consolidated scientific impact maintained since 2019. Briefings in Bioinformatics follow as the second most relevant source; despite its more recent publication history, it demonstrates high impact (h-index = 7) and a m-index of 1.167, signaling rapid growth within the field.

Journal of Molecular Sciences, shows high m-indices (1.5), indicating significant initial impact notwithstanding their currently moderate total citation counts. Conversely, Future Internet shows a substantial volume of citations but a more gradual growth in impact (m-index = 0.75). The remaining sources in the analyzed corpus report comparatively lower levels of impact and productivity.

Table 1. Impact of sources by bibliometric index (Bibliometrix)

Source	h_index	g_index	m_index	TC	NP	PY_start
IEEE Access	8	12	1.143	1333	12	2019
Briefings in Bioinformatics	7	11	1.167	349	11	2020
Expert Systems	4	5	1.0	42	5	2022
Journal of Chemical Information and Modeling	4	6	1.0	215	6	2022
Journal of Cheminformatics	4	6	1.333	161	6	2023
Artificial Intelligence Review	3	8	1.5	109	8	2024
Future Internet	3	5	0.75	467	5	2022
International Journal of Molecular Sciences	3	9	1.5	150	9	2024
Archives of Computational Methods in Engineering: State of the Art Reviews	2	2	0.4	27	2	2021
Artnodes	2	2	0.333	14	2	2020

Source: Authors' elaboration based on data from Impact of sources by bibliometric (Bibliometrix) (2025)

An analysis of the most relevant authors reveals a high concentration of research among a small group of researchers, characterized by significant co-authorship (Table 2). Zhang Y is the most prolific contributor with 8 articles and the highest fractional authorship (1.54), signifying a substantial individual. Wang X follows with 5 publications (0.93 fractional), while

Li C, Li Y, Wang Z, and Zhang J demonstrate moderate productivity (4 articles each) with fractional values ranging from 0.51 to 0.82.

Table 2. Most relevant authors by number of documents (Bibliometrix)

Author	Articles	Articles Fractionalized
Zhang Y	8	1.54
Wang X	5	0.93
Li C	4	0.51
Li Y	4	0.53
Wang Z	4	0.59
Zhang J	4	0.82
Chen L	3	0.41
Li X	3	0.44
Ali S	2	0.29
Bao T	2	0.40

Source: Authors' elaboration based on data from most relevant authors by number of documents by bibliometric (Bibliometrix) (2025)

4.4. Collaboration networks

The analysis of the author collaboration network (Figure 6) reveals a concentration of scientific influence among a few strongly interconnected key actors, alongside several isolated peripheral groups. Within the central core (red cluster), a high density of cooperation is observed, where Zhang Y serves as the primary central node (hub), directly linking to prolific researchers such as Wang X, Li Y, and Li C.

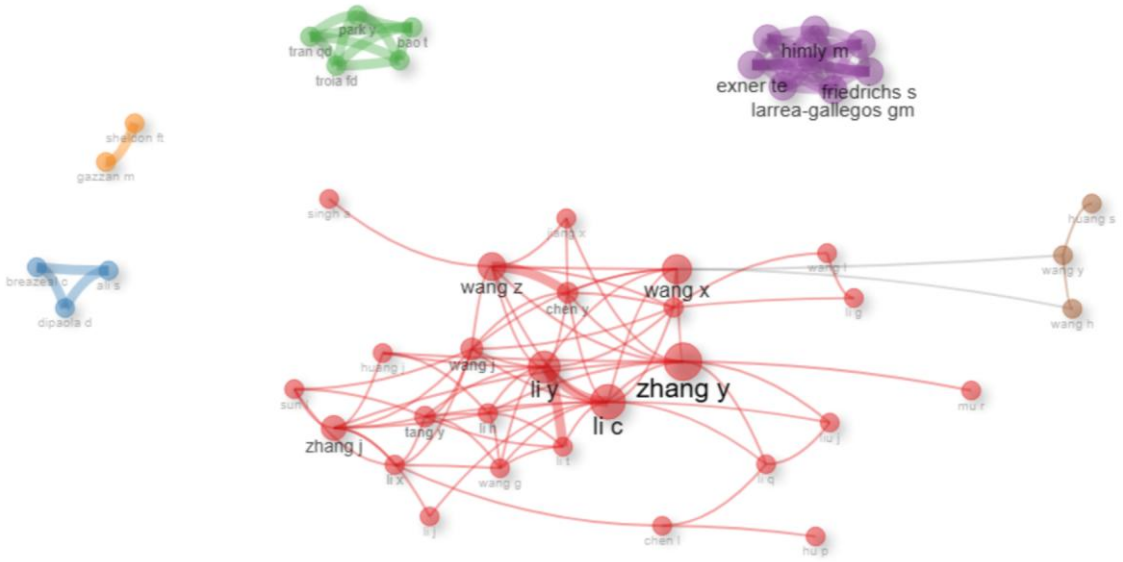

Fig 6. Collaboration network between authors
Source: output from Bibliometrix (2025).

4.5. Thematic Clusters and Architectures (QI2)

The keyword co-occurrence network (Figure 7) reveals a cohesive yet diverse research landscape, organized into two primary interconnected clusters. This structure indicates a mature field where methodological frameworks and technical implementations are deeply integrated.

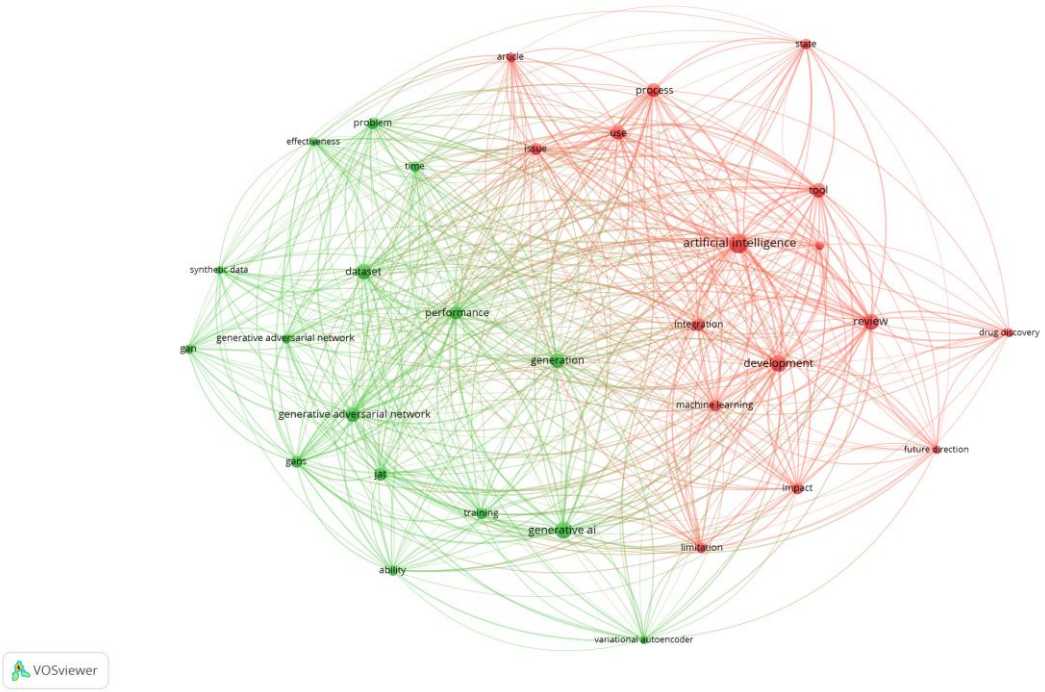

Fig 7. Thematic clustering of literature
Source: output from VOSviewer (2025).

Cluster red - Methodological and prospective: This grouping centers on foundational and transversal themes, such as “artificial intelligence”, “machine learning”, and “systematic reviews”.

Cluster green - Technical and generative: Focused on operational research, this cluster aggregates keywords related to generative architectures, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and synthetic data. The prevalence of terms like “training”, “datasets”, and “performance” reflects an experimental emphasis on model optimization and data synthesis.

The high link density between these clusters, bridged by pivotal nodes such as generation and integration, demonstrates that advancements in generative AI are intrinsically linked to broader machine learning developments. Overall, the network depicts a rapidly expanding domain where generative models serve as a central axis for recent scientific progress.

The thematic map in Figure 8 categorizes the conceptual structure of the studied domain into four quadrants, correlating relevance (centrality) with the degree of development (density) of the various topics.

- Motor Themes: These emerge as the current pillars of development, with significant prominence given to “Generative AI”, “Large Language Models (LLMs)”, and “Reinforcement Learning”. These themes exhibit high centrality and density, representing the leading edge of current research.
- Basic Themes: These represent transversal foundations, such as “Machine Learning”, “Artificial Intelligence”, and “Deep Learning”. Although fundamental to sustaining the field, they exhibit lower density, as they are generalist concepts that have become broadly integrated across literature.
- Niche Themes: Highly specialized areas, such as “diffusion models” and “drug design”, demonstrate isolated technical development. While technically sophisticated, these specific applications have not yet achieved global centrality within the broader research network.
- Emerging or Declining Themes: Topics such as “creativity” and “generative algorithms” suggest new research frontiers. Although these are currently in their nascent stages, they possess the potential to migrate toward quadrants of higher relevance as the field evolves.

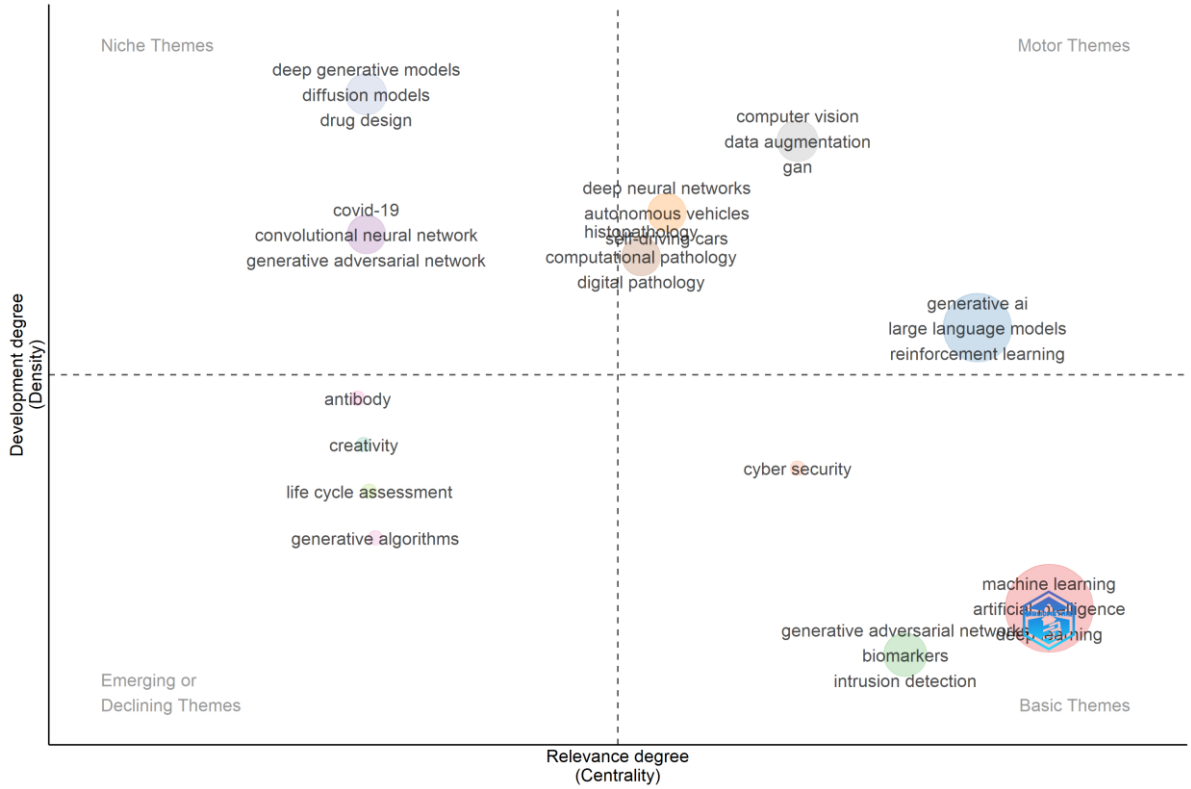

Fig 8. Thematic map – centrality versus density of themes
Source: output from Bibliometrix (2025).

In summary, the analysis reveals a research field in a state of advanced maturity and accelerated specialization. This is characterized by a definitive transition from general predictive models toward complex generative architectures, following a well-defined evolutionary trajectory.

5. Discussion

Bibliometric evidence suggests that GenAI in cybersecurity has moved from exploratory work (2018–2021) to rapid expansion after 2022, with a pronounced concentration of attention in 2024–2025 around LLM-centered automation. This acceleration is accompanied by an asymmetry between attack and defense: offensive capabilities (e.g., mutability, obfuscation, and social engineering scale) mature faster than defensive evaluation frameworks, particularly for adversarial robustness and functional validation of synthetic malware. Collaboration networks indicate an increase in international co-authorship, but also a concentration of production in a limited set of countries and institutions, in line with unequal access to

computing, data, and security test beds. In practice, the field would benefit from shared references and audit-ready evaluation protocols that integrate threat modeling, red teaming, and governance requirements (e.g., traceability, controllability, and accountability) for GenAI systems deployed in high-risk cyber contexts.

6. Conclusions

This review provides a PRISMA-guided bibliometric synthesis of research on General Artificial Intelligence (GenAI) in cybersecurity, concluding that the research field is growing rapidly, with the number of publications increasing sharply after 2022 and focusing on a limited set of stakeholders.

Based on the formulated research questions, the analysis of the results allows a set of conclusions to be drawn. Regarding RQ1, scientific production in the field of GenAI applied to cybersecurity shows a marked increase from 2022 onward, concentrated in a limited number of publication venues (notably IEEE Access) and among a small group of highly productive authors (with prominence of Zhang Y.).

Regarding RQ2, the literature reveals the predominance of well-defined thematic clusters and specific GenAI architectures. Dominant themes cluster around (i) defensive applications such as intrusion detection systems (IDS), cyber threat intelligence (CTI), and synthetic data generation; (ii) mutability, offensive evasion, and adversarial attacks; and (iii) IoT and edge security under resource-constrained environments.

Concerning RQ3, the distribution of the literature between defensive applications and research oriented towards attack or evasion indicates a predominance of defensive approaches in terms of the volume of empirical studies. Nevertheless, offensive research remains persistent and dynamic, particularly in contexts where large language models (LLMs) enable rapid and automated code transformations, thereby facilitating novel attack and evasion techniques.

Finally, the analysis associated with RQ4 highlights a set of significant limitations and research gaps that open avenues for future work. Key gaps include auditability and governance mechanisms for GenAI systems, the adversarial robustness of defensive models, and the lack of standardized protocols to validate the functional fidelity of synthetic malware.

Despite its contributions, this study presents certain limitations. Primarily, the data collection was restricted to a single database (The Lens), and the dataset for 2025 remains incomplete due to the timing of the analysis. Additionally, as the bibliometric mapping is strictly quantitative, further qualitative research is required to provide a more nuanced understanding of this field of study. It is also suggested that future research develop: i) Standardized references for mutable malware and GenAI-assisted attacks; ii) Adversely robust IDS pipelines, validated under realistic attacker adaptation; iii) Governance and audit mechanisms that enable traceable, controllable, and accountable deployment of GenAI in security operations; iv) Plans for hybrid defensive architectures. These architectures should integrate generative AI alongside formal verification and symbolic validation techniques, which would help reinforce systems to combat complex and constantly changing attacks. These architectural approaches really need to integrate generative AI in a continuous manner.

The study demonstrates that the results of this bibliometric study show that generative AI is indeed transforming cybersecurity, even if it is not yet fully complete. Some important audit frameworks are indeed necessary, in addition to risk assessment measures that are attentive to model changes and governance structures that integrate clarity and explainability with accountability. This analysis demonstrates that any future research focused on the technical effectiveness and reliability, resilience, and overall sustainability of AI-driven cybersecurity infrastructure is indeed vital and important.

References

Acosta-Bermejo, R., Terrazas-Chavez, J. A., & Aguirre-Anaya, E. (2025). Automated Malware Source Code Generation via Uncensored LLMs and Adversarial Evasion of Censored Model. *Applied Sciences (Switzerland)*, *15*(17). https://doi.org/10.3390/app15179252

Balasubramanian, P., Liyana, S., Sankaran, H., Sivaramakrishnan, S., Pusuluri, S., Pirttikangas, S., & Peltonen, E. (2025). Generative AI for cyber threat intelligence: applications, challenges, and analysis of real-world case studies. *Artificial Intelligence Review*, *58*(11). https://doi.org/10.1007/s10462-025-11338-z

Dunmore, A., Jang-Jaccard, J., Sabrina, F., & Kwak, J. (2023). A Comprehensive Survey of Generative Adversarial Networks (GANs) in Cybersecurity Intrusion Detection. *IEEE Access*, *11*, 76071–76094. https://doi.org/10.1109/ACCESS.2023.3296707

Gaber, M. G., Ahmed, M., & Janicke, H. (2024). Malware Detection with Artificial Intelligence: A Systematic Literature Review. *ACM Computing Surveys*, *56*(6). https://doi.org/10.1145/3638552

Gazzan, M., Alobaywi, B., Almutairi, M., & Sheldon, F. T. (2025). A Deep Learning Framework for Enhanced Detection of Polymorphic Ransomware. In *Future Internet* (Vol. 17, Issue 7). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/fi17070311

Guo, Y. (2023). A Survey of Machine Learning-Based Zero-Day Attack Detection: Challenges and Future Directions. In *Computer Communications* (Vol. 198, pp. 175–185). Elsevier B.V. https://doi.org/10.1016/j.comcom.2022.11.001

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. In *IEEE Access* (Vol. 11, pp. 80218–80245). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2023.3300381

Ibrar, W., Mahmood, D., Al-Shamayleh, A. S., Ahmed, G., Alharthi, S. Z., & Akhunzada, A. (2025). Generative AI: a double-edged sword in the cyber threat landscape. *Artificial Intelligence Review*, *58*(9). https://doi.org/10.1007/s10462-025-11285-9

Javaheri, D., Lalbakhsh, P., & Hosseinzadeh, M. (2021). A Novel Method for Detecting Future Generations of Targeted and Metamorphic Malware Based on Genetic Algorithm. *IEEE Access*, *9*, 69951–69970. https://doi.org/10.1109/ACCESS.2021.3077295

Kumar, D., Pawar, P. P., Addula, S. R., Meesala, M. K., Oni, O., Cheema, Q. N., Haq, A. U., & Sajja, G. S. (2025). AI-Powered Security for IoT Ecosystems: A Hybrid Deep Learning Approach to Anomaly Detection. *Journal of Cybersecurity and Privacy*, *5*(4), 90. https://doi.org/10.3390/jcp5040090

Labaca-Castro, R. (2023). Machine Learning under Malware Attack. In *Machine Learning under Malware Attack*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-40442-0

Liu, Y., Huang, J., Li, Y., Wang, D., & Xiao, B. (2025). Generative AI model privacy: a survey. *Artificial Intelligence Review*, *58*(1). https://doi.org/10.1007/s10462-024-11024-6

Liu, Y., Huang, J., Li, Y., Wang, D., & Xiao, B. (2025). Generative AI model privacy: a survey. *Artificial Intelligence Review*, *58*(1). https://doi.org/10.1007/s10462-024-11024-6

Mahmoudi, I., Boubiche, D. E., Athmani, S., Toral-Cruz, H., & Chan-Puc, F. I. (2025). Toward Generative AI-Based Intrusion Detection Systems for the Internet of Vehicles (IoV). In *Future Internet* (Vol. 17, Issue 7). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/fi17070310

Pei, G., Ma, K., Eece, ucasaccn, Beijing, U., Dongpeng Zhang, C., Xu, Q., Huang, Q., Zhang, D., & Sun, C. (n.d.). A Unified Framework for Stealthy Adversarial Generation via Latent Optimization and Transferability Enhancement. In *Proceedings of the 33nd ACM*

International Conference on Multimedia (MM '25), October 27â•fiOctober 31, 2025, Dublin, lreland (Vol. 1). ACM.

Radanliev, P. (2025). Collaborative penetration testing suite for emerging generative AI algorithms. *Applied Intelligence*, *55*(16). https://doi.org/10.1007/s10489-025-06908-1

Yazdani, S., Singh, A., Saxena, N., Wang, Z., Palikhe, A., Pan, D., Pal, U., Yang, J., & Zhang, W. (2025). Generative AI in depth: A survey of recent advances, model variants, and real-world applications. *Journal of Big Data*, *12*(1). https://doi.org/10.1186/s40537-025-01247-x