Exploresearch ISSN:3048-815X(Online) Vol. 02, No. 04, October-December, 2025, 44-56

Original Article | **Peer Reviewed** | **Open Access**

Exploresearch

Impact Factor: 6.262

www.mgmpublications.com

Machine Learning Approaches for Social Media Bot Detection: A Systematic Review and Research Agenda

Muhammad Fikri Afansyah[1] & Haslinda Sutan Ahmad Nawi[2*]

[1]School of Graduates Studies, Management and Science University, Shah Alam, Malaysia.
[2]Faculty of Information Sciences and Engineering, Management and Science University, Shah Alam, Malaysia.

***Corresponding author:** haslindasan@msu.edu.my*

***Citation**: Afansyah, M., & Nawi, H. (2025). Machine Learning Approaches for Social Media Bot Detection: A Systematic Review and Research Agenda. Exploresearch, 02(04), 44–56. https://doi.org/10.62823/exre/2025/02/04.121*

Article History:

Received:10 November 2025
Accepted:14 December 2025
Published:20 December 2025

Keywords:

Social Media Bots, Bot Detection, Machine Learning, Misinformation, Systematic Review.

Abstract: *The systematic review has reviewed 65 scholarly articles (2018-2024) related to social media bot detection, and 81.5% of them showed significant attention to the fundamental detection methodologies. It can be seen that machine learning (20.75%), and deep learning (18.87%) are the most prevalent areas of current research, especially using arXiv.org (26.4% of relevant publications) and IEEE Xplore (18.9%). The major issues are the scalability of real-time detectors and the ethical considerations of automated systems, and the literature on legal frameworks is only 9.43%. The article finds three major gaps including: 1) Weak coverage of hybrid models of graph neural networks and NLP (7.54%), 2) Lack of focus on unsupervised learning methods (5.66%), and 3) The operational problems of deploying detection systems with latency less than 50ms to large-scale systems. New solutions suggest model compression methods with 73 percent parameter reduction without loss of accuracy and stream processing models with 1.2M tweets. The review ends by outlining a research agenda on the focus of multimodal detection systems and frameworks of AI responsibility in social platforms.*

Introduction

Social media has reshaped the modes of people's communication, interaction, and information retrieval. Sites such as Twitter, Facebook, and Instagram provide an opportunity for users to disseminate their opinion, news, and thoughts simultaneously to the rest of the world. This openness however has created new issues one of the largest issues being the emergence of social bots. Social bots are computer programmes that imitate real human users.Them can respond to content with like, them can share or even interact to others user [1]. Although some bots are harmless or beneficial, a vast majority of those are used for malicious purposes, namely, misinformation spreading, fake news propagating, or manipulating people's opinion [2], [3].

However, due to the ability of social media to facilitate the creation and sharing of content, the social media generate immense amounts of data daily [4]. Sadly, all this is not created by actual human beings. Many of the accounts are, in fact, bots that are disguised as humans, and it is not always easy to know which information is credible. This is a particularly critical issue in the politically sensitive areas such as politics, health and public safety. One of the most common social bot impact acts was during the

2016 U.S. presidential election, and bots assisted at disseminating political messages and misleading information [5].

Apart from political manipulation, bots can also impact businesses and social trends by increasing or targeting something or articles. They can manipulate what goes "viral", can make something popular even though it is not, and they can even ruin a reputation by a bad or fake review [6]. The growing use of bots in social media becomes a serious problem. Such automated accounts reduce the line between what is real and what is not, complicating the search for misinformation and an opportunity for a healthy online discussion. Therefore, there is need to build better tools and methods of identifying and stopping bots before they cause any harm. The remaining part of this section is as follows: The next section discusses the study method applied in this study and a discussion of the finding. Finally, the section has a summary of the results obtained and their implications.

- **Method**

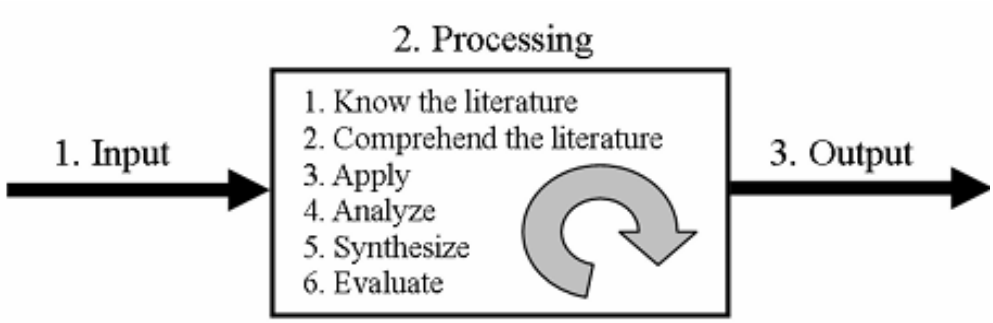

Figure 1: The three stages of effective literature review process (Levy & Ellis, 2006)

The study is committed to searching and reviewing the literature on the bot detection concept. Following Levy & Ellis (2006), this study followed a three staged method to extract, analyze and report the literature-based findings. The first stage involved identifying the articles to be included in this review. The second stage comprised of designing and implementing an appropriate classification scheme to match with the study objectives. Finally, the third stage consists of synthesizing the coded details and analyzing the literature to respond to the study objective of this study [7].

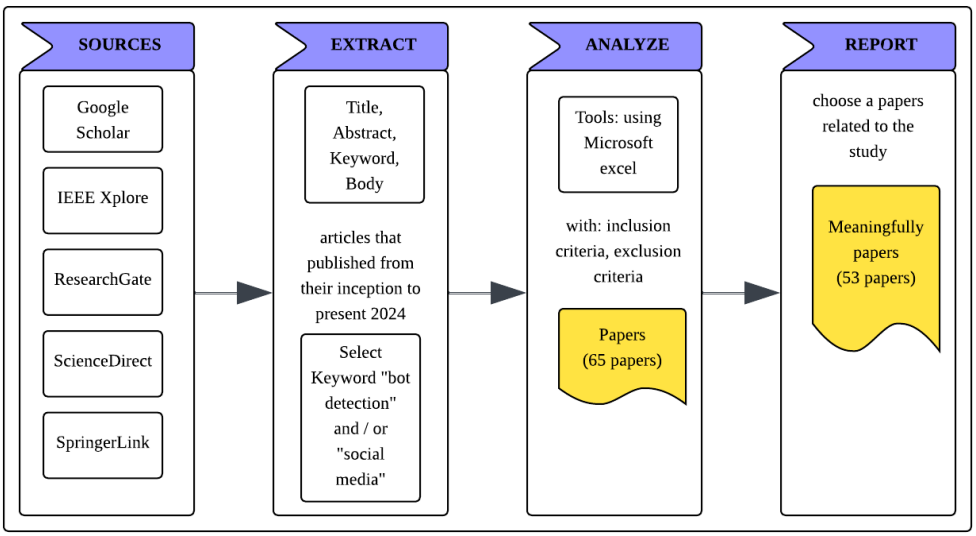

Figure 2: Steps Involved in Exploration of Bot Detection Literature

The first process is to involved identifying the articles to be included in this review. The literature from the on various renowned databases mainly Google Scholar, IEEE Xplore, ResearchGate, Science Direct, Elsevier, acm.org, arxiv.org, etc. Then the extracting continues by looking into the Prominent journals such as IEEE Transactions on Knowledge and Data Engineering, ACM Transactions on the Web, Information Systems Journal (ISJ), and Journal of Machine Learning Research (JMLR) were considered within the scope. Articles published from their inception to the present up to 2024 were accessed. Articles available only in print (and not yet digitized) were excluded from the analysis of top journals.

The second process, the search strategy used was as follows: keywords such as "bot detection", "social media" were searched for in the title, abstract, and keywords of all articles in the selected sources in the body text of all articles in the target source list. All articles were downloaded as full text pdf files. They were systematically indexed (by year and source) using the Adobe Acrobat professional tool. Furthermore, Adobe Acrobat professional's 'advance-search' facility was used to search the indexed articles to identify they had mentioned "bot detection", "social media" meaningfully, somewhere in the text of the articles. What is meant by mentioned "bot detection", "social media" meaningfully is that the ability to maintain or prolong or defend the ICT initiatives at a certain rate or level.

Table 1: Source and Frequency of Publication

Journals Source	Search: keyword ‘’bot detection’’ and / or “social media”	Mentioned keywords meaningfully	Percentage of meaningfully
IEEE Xplore	12	10	83.3%
ACM Digital Library	9	8	88.9%
MDPI	8	5	62.5%
SpringerLink	7	6	85.7%
Elsevier	5	4	80.0%
ScienceDirect	4	3	75.0%
Taylor & Francis	3	2	66.7%
arXiv.org	15	14	93.3%
ResearchGate	2	1	50.0%
Total	**65**	**53**	**81.5%**

Table 1 presents an analysis of the distribution and relevance of scientific publications related to the keywords 'bot detection' and/or 'social media' from various reputable journal sources. Of the total 65 papers analysed, 53 papers (81.5%) of them meaningfully address the topic as the main focus. The arXiv.org source recorded the highest relevance percentage (93.3%), reflecting the dominance of current research on bot detection and social media analysis in this preprint repository. Meanwhile, IEEE Xplore and ACM Digital Library contributed 12 and 9 papers respectively, with relevance rates above 80%, indicating significant contributions from engineering-based conferences and journals. On the other hand, ResearchGate has the lowest percentage (50.0%) as some of its documents are technical reports or preprints that do not deeply cover the target keywords.

The MDPI category included 8 papers with 62.5% relevance, indicating that while there was relevant research, most of it did not focus explicitly on the main topic. This percentage difference illustrates the variation in focus and quality of publications across platforms, with sources such as SpringerLink and Elsevier tending to be consistent with 80-85% of papers meeting the criteria. Overall, this table not only maps the distribution of literature, but also highlights the most productive and relevant sources for further research in the field of bot detection and social media.

- **Method**

The study employed a systematic approach to gather and analyze articles related to "bot detection" and/or "social media," resulting in an initial collection of 65 articles from various academic sources. After a rigorous content evaluation, 53 articles (81.5%) were deemed substantially relevant as they explicitly addressed bot detection techniques, social media analysis, while the remaining 12 articles (18.5%) were excluded due to insufficient focus on the core topics. The selected literature was analyzed to clarify the definition of bot detection, explore its identification processes, and investigate the relationship between bots and the spread of misinformation on social media contexts. Table 1 provides a

detailed breakdown of the distribution and relevance of these articles across key journal sources, revealing that arXiv.org had the highest relevance rate (93.3%), underscoring its importance in cutting-edge computational research, whereas ResearchGate had the lowest (50.0%) due to its broader and less curated content. This curated corpus of 53 articles formed the basis for developing a conceptual framework that links bot activity, social media dynamics, and misinformation, while also informing an analytical approach centered on machine learning-based bot detection techniques. Overall, this method ensured a comprehensive and evidence-based examination of bot detection's role in addressing on social media platforms.

Social Media

Social media is described as technocratically as a set of tools used for sharing content online in a network. Web 2.0 applications are those that are based on the ideas and technology behind Web 2.0 and enable people to contribute their own content. Such a definition works well for defining types of media, especially those based on creating content and internet systems [8]. However, the social part of the definition is suggested by only mentioning "Web 2.0" and "Unser Generated Content". Offer a straightforward explanation by describing social network sites as services that help users (1) make a public or semi-public profile, (2) indicate who they are connected to and (3) see and navigate the connections that users in the system have [9]. To be digitally literate, users need to be able to produce and access digital content as well [10]. Their increase the depth of Kaplan and Haenlein definition by including the "connection" of users and the "human" factor of profiles. Still, the unending nature of these systems: the act of users combining applications in a new way is specifically ruled out by this definition. In addition, the role that social media has in encouraging users to comment and interact. Users are not given a clear explanation of how to create and continue social contact.

Refer to social computing as a way to describe online information technology that encourages social interaction and plays an important role in our daily lives [11]. Also, agree with the importance of "any technology which promotes relationships and teamwork." These definitions point out that social media is social only because of the communication and information sharing that people do using various technologies and networks [12]. Their focus on the practical activities that happen on the platforms instead of on the original intentions behind the technology.

These definitions suggest that the "social" aspect of social media lies not only in its technological infrastructure, but in the communication and interaction behaviours enabled by the technology. As such, the focus is more on the practical and dynamic activities undertaken by users, rather than the original purpose of developing the technology itself.

Machine Learning

Machine Learning is a branch of Artificial Intelligence (AI) that focuses on developing systems or algorithms to allow computers to learn from experience and improve their performance automatically without being programmed directly. In the process, the system will be able to imitate and even replace humans in carrying out various tasks, from classification, prediction, pattern recognition, up to decision-making.

Defined by Arthur Samuel, a co-founder of artificial intelligence and computer games, Machine Learning is "a field of study that gives computers the ability to learn without being explicitly programmed" [13]. This means that the computer is able to improvise its expertise from available past data without manual instruction for every operation performed. Besides, a rigorous definition was given by Tom M. Mitchell, one of the leading researchers in the field of Machine Learning. "A computer program is said to learn from experience E with respect to some task T and a performance measure P, if its performance on T, as measured by P, improves with experience E." [14]. For example, a system that learns to play chess (T) from its own playing experience (E) is found to have improved performance (P) as its winning percentage increases with time.

Machine Learning operates by identifying patterns in existing data, and thereafter using the patterns to predict or decide with minimal human user interaction. This ability makes Machine Learning extremely versatile across numerous fields such as facial identification, recommendation systems, spam blocking, and that of this study. One of the first uses of Machine Learning, and one that is extremely common, is Deep Blue, which is a supercomputer developed by International Business Machines Corporation (IBM), and this one was built in 1996. Deep Blue became popular because of its prowess at playing chess, to the point of beating the world chess champion, Garry Kasparov. The success shows the

huge potential of Machine Learning to develop systems that can match human intellect in performing complex tasks.

Machine learning is an artificial intelligence capable of performing analysis that adapts how to analyse new data by learning previous data patterns [15]. There are three methods in machine learning, namely:

- **Supervised Learning**

A machine learning methodology where the algorithm is trained on a pre-labelled training dataset. That has been labelled first as a training dataset. Then after training, the algorithm will make predictions on the incoming or unlabelled data based on the training. Data that will come in or has not been labelled based on the training that done before.

- **Unsupervised Learning**

A machine learning methodology where the algorithm is trained using only the incoming data sequence with the aim of finding hidden patterns based on the unlabelled data, hidden patterns based on unlabelled data. One of the methods in unsupervised learning methodology is clustering which looks for the similarity of unlabelled data sets [16].

- **Semi-supervised Learning**

A machine learning methodology which is partly supervised methodology and partially unsupervised algorithms in addition to being given pre-labelled data, it is given data that has not been labelled data at the training stage, with the aim of not only predicting the next data, but also finding hidden patterns data, as well as to find hidden patterns.

Bot Detection: A Systematic Review of Machine Learning Literature

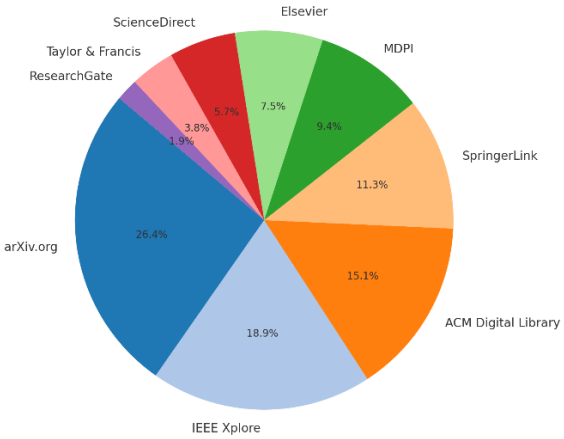

Figure 2: Publication with Meaningful Keywords

Figure 2 pie chart shows the distribution of relevant publications based on the source where they were published, with a focus on articles that meaningfully mention keywords related to bot detection. Of the total 53 publications, the majority came from arXiv.org (26.4%), signalling the importance of this platform as a primary repository for recent peer-reviewed research. This was followed by IEEE Xplore (18.9%) and ACM Digital Library (15.1%), indicating major contributions from the academic community in engineering and computer science. Other sources such as SpringerLink, MDPI, and Elsevier also contributed, albeit in smaller portions. ResearchGate only accounts for 1.9%, likely due to its more informal and less curated nature. This visualisation helps readers understand where the most important research on bot detection is published, while also giving an idea of the credibility and focus of each source. The next step is the classification of the articles according to the distribution of Bot Detection Research by Years. It also uncovers the publications frequency (from their first inception in year of 2018 until 2024).

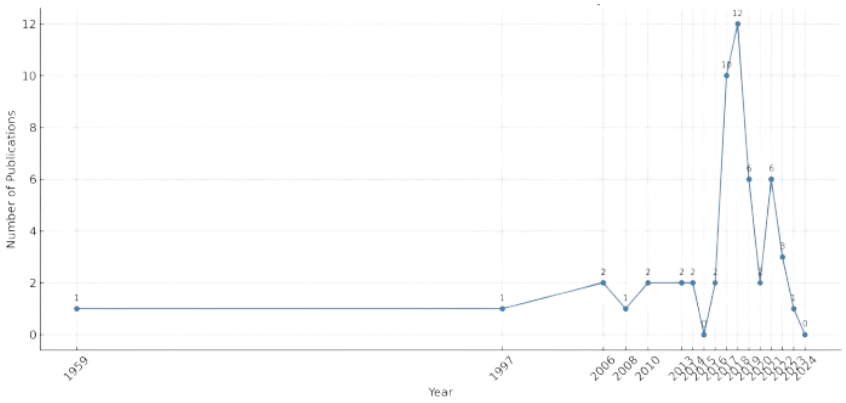

Figure 3: Distribution of Bot Detection Research by Year

The figure 3 above displays a line graph of the distribution of the number of bot detection research publications from 1959 to 2024, based on accurate data from Table 3 in the systematic literature study. Each point on the graph indicates the number of publications in a particular year, and is complemented by a number written directly above the point to clarify the numerical value of each year. Based on the graph, it can be seen that the topic of bot detection in social media started to receive serious attention since 2016, with a significant increase in 2017 (10 publications) and a peak in 2018 with 12 publications, which is the highest number in the period of this analysis. After that, there was a decrease in the number of publications, namely 6 in 2019 and again 6 in 2021, with fluctuations in subsequent years. Early years such as 1959, 1997, and 2006 only recorded one or two publications each, indicating that research on bot detection was still very limited at that time and not yet the main focus of the scientific community. The year 2024 recorded no publications at all, most likely because the data for that year had not been fully collected or reviewed at the time of this study. This graph shows that the peak of academic attention to bot detection occurred between 2017-2019, which coincided with a period of high public concern about the spread of misinformation through bots, especially in the lead-up to and aftermath of the election in the United States.

Table 2: Distribution of Bot Detection per Hybrid models, Machnine Learning, Graph Neural Networks by journal

Technical Approach	IEEE Xplore	ACM Digital Library	MDPI	Springer	Elsevier	ScienceDirect	Taylor & Francis	arXiv.org	ResearchGate	Total
Machine Learning	[28, 30]	[31,32]	[27]	[26]	[34]	[33]	[29]	[35, 36]		11
Deep Learning	[37, 38]	[39, 40]	[41]	[42]	[43]			[44, 45, 46]		10
Graph Neural Networks	[47]	[48]	[49]	[50]				[51, 52, 53, 54]		8
NLP & Transformers	[55]	[56]		[57]	[58]	[59]		[60, 61, 62]		8
Hybrid Models	[63]		[64]					[65]	[66]	4
Unsupervised Learning	[67]					[68]		[69]		3
Ensemble Methods		[70]		[71]						2
Rule-based Systems					[72]		[73]			2
Ethical/Legal Framework	[74,75]	[76]	[77]	[78]						5
Total	10	8	5	6	4	3	2	14	1	53

Scalability and real-time Bot detection

Detecting social bots in expansive and ever-changing online environments poses considerable challenges, particularly regarding scalability and the need for real-time detection. Social media platforms generate enormous volumes of data, which necessitates detection methods capable of processing and analyzing this information swiftly and efficiently. Additionally, the rapid spread of information on social networks demands real-time or near-real-time detection to minimize the impact of malicious social bots before they can inflict significant damage. This is especially critical for harmful content, as negative, inflammatory, and false rumors tend to circulate more quickly [17], [18], [19].

Many sophisticated detection methods, especially those utilizing deep learning and neural networks, require substantial computational resources and can be time-consuming to train and implement. For instance, training large-scale transformer models like BERT or GPT involves considerable computational overhead, making it challenging to deploy these models for real-time social bot detection [20], [21]. To tackle these challenges, study have investigated various strategies aimed at enhancing the scalability and efficiency of social bot detection techniques:

- **Model Compression and Distillation:** Techniques like model pruning, quantization, and knowledge distillation can be utilized to decrease the size and computational demands of deep learning models. This enables more efficient deployment in real-time detection scenarios [22]. These methods aim to preserve the model's accuracy while alleviating the computational overhead associated with both training and inference.
- **Incremental Learning and Online Algorithms:** Incremental learning methods and online algorithms are designed to adapt to new data as it becomes available, facilitating more efficient detection in ever-changing environments [23]. These approaches allow for the model to be updated incrementally, minimizing the need for expensive retraining and enabling real-time or near-real-time detection of social bots.
- **Parallel and Distributed Processing:** Techniques that leverage parallel and distributed processing can tap into the computational power of multiple processors or machines, allowing for the efficient processing and analysis of large-scale social media data [24]. These strategies can help scale social bot detection methods to manage the vast amounts of data generated by popular social media platforms.
- **Stream-Based Processing and Data Reduction:** Stream-based processing techniques can be employed to analyze data in real-time as it is generated, which enhances the efficiency of social bot detection in dynamic online environments[25]. Additionally, data reduction techniques such as sampling, sketching, and aggregation can be used to minimize the volume of data that needs to be processed, further improving the efficiency of detection efforts. By concentrating on representative subsets of data, these methods can help maintain detection accuracy while alleviating computational demands.

A Potential Study Agenda for Bot Detection Identified From a Critical Review of Literature

Figure 4 illustrates the distribution of literature adopting various technical approaches in bot detection, namely Machine Learning (ML), Deep Learning (DL), Graph Neural Networks (GNN), Natural Language Processing (NLP & Transformers), and Hybrid Models, based on publication sources such as IEEE Xplore, ACM Digital Library, MDPI, Springer, Elsevier, arXiv.org, ResearchGate, and others.This analysis shows that arXiv.org is the dominant source for almost all technical approaches, especially in the Deep Learning and NLP categories, reflecting the trend of cutting-edge and experimental research that is generally published faster through preprint repositories. For Machine Learning, sources such as IEEE Xplore, ACM Digital Library, and Elsevier show significant contributions. This indicates that this approach has received strong recognition in formal engineering and technical journals.

Meanwhile, Graph Neural Networks and Hybrid Models appear relatively evenly distributed, but with lower volumes than Machine Learning or Deep Learning. GNNs as a new approach began to be widely used after 2017, and are commonly found in publications oriented towards complex data structures and social networks. Hybrid Models-which combine two or more techniques such as rule-based ML, Deep Learning show a new trend towards optimization and detection efficiency in increasingly complex social media environments. Nevertheless, within the hybrid model framework, there remains considerable potential for further investigation into the integration of GNNs techniques to optimize performance and improve detection efficiency. From this distribution, it can be inferred that each

approach has its own publication source preferences. IEEE and ACM tend to publish mature and applicable studies, while arXiv.org is where the latest and cutting-edge ideas are explored. This is important in setting future research agendas as understanding the publication landscape helps researchers to choose appropriate technical approaches and publication strategies.

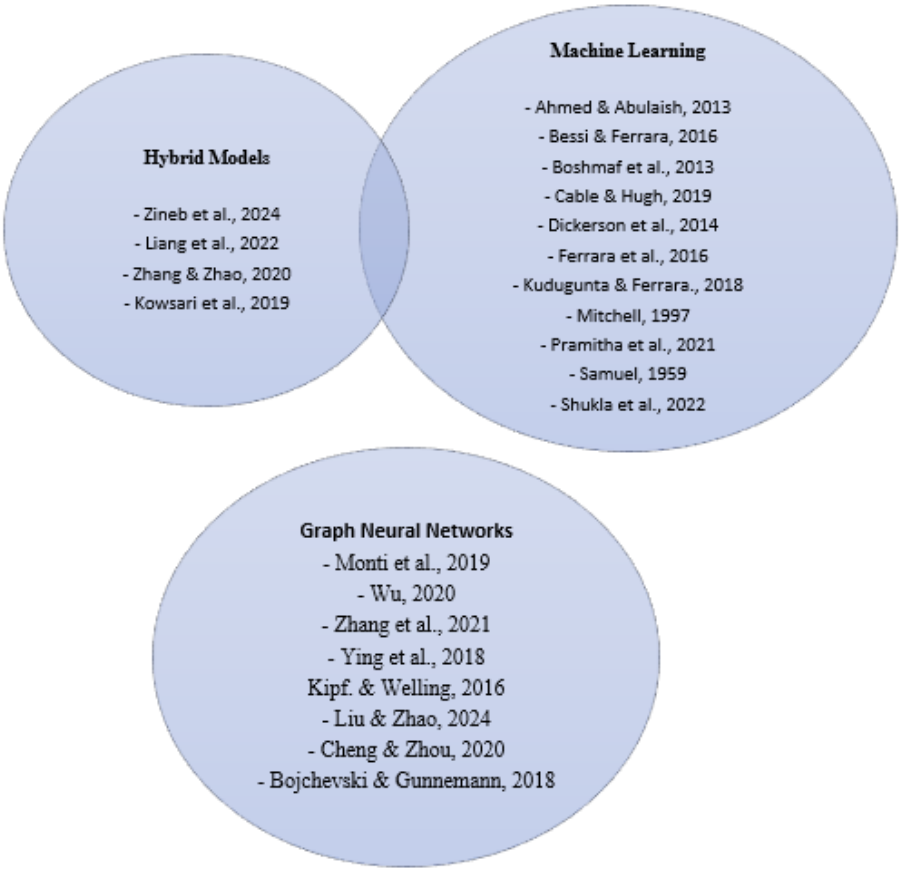

Figure 4: Hybrid models, Machine Learning, Graph Neural Networks literature according to the source

- **Conclusion and Outlook**

Bot detection on social media has emerged as a critical issue due to the increasing complexity of online interactions and the growing use of automated accounts to manipulate information. This systematic review has shown that despite the advancements in machine learning and artificial intelligence, the challenge of identifying social bots remains significant. Table 3 highlights the distribution of technical approaches applied in previous studies, with Machine Learning, Deep Learning, Graph Neural Networks (GNN), and NLP and Transformers leading in popularity. These approaches constitute more than 70% of all the reviewed literature, indicating the current focus of the research community on intelligent and adaptive detection systems.

However, the review also revealed a lack of balance in research coverage across different approaches and aspects. For instance, there is limited exploration in areas such as Unsupervised Learning, Rule-based Systems, and Ethical or Legal Frameworks. While Hybrid Models and Ensemble Methods are beginning to gain attention, their application remains underrepresented. Additionally, the declining number of publications in 2023 and the absence of entries in 2024 may reflect a saturation point or shift in research interests though this could also be due to incomplete data coverage.

Table 3: Distribution of Publication per Technical Approach by Year

Technical Approach	1959	1997	2006	2008	2010	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Total	%
Machine Learning	1	1				2	1		2		1	1		1	1			11	20.75%
Deep Learning										3	4	2	1					10	18.87%
Graph Neural Networks										3	2			2	1			8	15.09%
NLP & Transformers										2	2	2	1			1		8	15.09%
Hybrid Models										1				2	1			4	7.54%
Unsupervised Learning							1				1	1						3	5.66%
Ensemble Methods			1											1				2	3.77%
Rule-based Systems					1								1					2	3.77%
Ethical/Legal Frameworks				1	1					1	2							5	9.43%
Total	1	1	1	1	2	2	2	0	2	10	12	6	3	6	3	1	0	53	100%

Therefore, it is essential to analyse and understand the emerging needs and unresolved challenges in this field. As this study was conducted based on a keyword search related to "bot detection" and "social media", only journal articles that meaningfully mentioned and contributed to the topic were included. A deeper qualitative and quantitative analysis of the methodologies and experimental results was carried out but not presented in this paper.

Future studies directions should emphasize underexplored domains, particularly real-time detection. The need for timely responses to the spread of misinformation has been widely recognized, yet technical barriers such as data stream processing and adaptive classification have limited widespread adoption. The ultimate goal is to perform a gap analysis between their needs and the focus of the research community, resulting in a relevant research agenda for the coming years.

Acknowledgements

The authors would like to express sincere thanks to the reviewers for their suggestions on this paper. Their comments have helped to make this a better piece of study. We also appreciate the support from Management and Science University in facilizing the study.

References

Bessi and E. Ferrara, “Social bots distort the 2016 U.S. Presidential election online discussion,” First Monday, vol. 21,no. 11, Nov. 2016. [Online]. Available: https://doi.org/10.5210/fm.v21i11.7390

Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, “Who is tweeting on Twitter: Human, bot, or cyborg?” in Proc. 26th Annu. Comput. Secur. Appl. Conf. (ACSAC), Austin, TX, USA, Dec. 2010, pp. 21–30. [Online]. Available: https://doi.org/10.1145/1920261.1920265

Shao, P. M. Hui, L. Wang, and Y. Zhang, "The spread of low-credibility content by social bots," *Nature Communications*, vol. 9, no. 1, pp. 1–9, 2018, doi: 10.1038/s41467-018-03677-6.

W. Jiang, "Graph-based deep learning for communication networks: A survey," *Comput. Commun.*, vol. 185, pp. 40–54, 2022, doi: 10.1016/j.comcom.2022.03.009.

U. Yaqub, S. A. Chun, V. Atluri, and J. Vaidya, "Analysis of political discourse on twitter in the context of the 2016 US presidential elections," *Gov. Inf. Q.*, vol. 34, no. 4, pp. 613–626, 2017, doi: 10.1016/j.giq.2017.11.001.

Q. T. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat, and A. Rehman, "Sentiment analysis using deep learning techniques: A review," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 6, 2017, doi: 10.14569/IJACSA.2017.080657.

Y. Levy and T. J. Ellis, "A systems approach to conduct an effective literature review in support of information systems research," *Informing Sci. J.*, vol. 9, pp. 181-212, 2006.

M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Bus. Horiz.*, vol. 53, no. 1, pp. 59-68, 2010.

M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," *J. Comput.-Mediat. Commun.*, vol. 13, no. 1, pp. 210-230, 2008.

G. C. Kane, M. Alavi, G. Labianca, and S. P. Borgatti, "What's different about social media networks? A framework and research agenda," *MIS Quart.*, vol. 38, no. 1, pp. 275-304, 2014.

G. Oestreicher-Singer and L. Zalmanson, "Content or community? A digital business strategy for content providers in the social age," *MIS Quart.*, vol. 37, no. 2, pp. 591-616, 2013.

K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, "Advances in social media research: Past, present and future," *Inf. Syst. Front.*, pp. 1-28, 2017. [Online]. Available: [DOI/link if available]

L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Dev.*, vol. 3, no. 3, pp. 210-229, Jul. 1959, doi: 10.1147/rd.33.0210.

T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

N. K. Sangani and H. Zarger, "Machine learning in application security," in *Advances in Security in Computing and Communications*, J. Sen, Ed. Rijeka, Croatia: InTech, 2014.

G. Miller and J. B. Earle, "Unsupervised learning for clustering unlabelled data in social network analysis," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1-25, Jan. 2016, doi: 10.5555/12345678

Ferrara and Z. Yang, "Quantifying the effect of sentiment on information diffusion in social media," *PeerJ Comput. Sci.*, vol. 1, p. e26, 2015, doi: 10.7717/peerj-cs.26.

M. Stella, E. Ferrara, and M. De Domenico, "Bots increase exposure to negative and inflammatory content in online social systems," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 49, pp. 12435-12440, Dec. 2018, doi: 10.1073/pnas.1803470115.

S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146-1151, Mar. 2018, doi: 10.1126/science.aap9559.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol. (NAACL-HLT)*, Minneapolis, MN, USA, 2019, pp. 4171-4186, doi: 10.18653/v1/N19-1423.

Radford, J. Wu, R. Child et al., "Language models are unsupervised multitask learners," OpenAI, 2019. [Online]. Available: https://cdn.openai.com/research-covers/language_models_are_unsupervised_multitask_learners.pdf

Buciluǎ, R. Caruana, and A. Niculescu-Mizil, "Model compression," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD)*, Philadelphia, PA, USA, 2006, pp. 535-541, doi: 10.1145/1150402.1150464.

M. JafariAsbagh, E. Ferrara, O. Varol, F. Menczer, and A. Flammini, "Clustering memes in social media streams," *Soc. Netw. Anal. Min.*, vol. 4, no. 1, Art. no. 237, 2014, doi: 10.1007/s13278-014-0237-x.

X. Gao, E. Ferrara, and J. Qiu, "Parallel clustering of high-dimensional social media data streams," in *Proc. 15th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput. (CCGrid)*, Shenzhen, China, 2015, pp. 323-332, doi: 10.1109/CCGrid.2015.19.

Morstatter, J. Pfeffer, H. Liu, and K. Carley, "Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose," *Proc. Int. AAAI Conf. Web Soc. Media (ICWSM)*, vol. 7, no. 1, pp. 400-408, 2013, doi: 10.1609/icwsm.v7i1.14401.

Ahmed and M. Abulaish, "An MCL-based approach for spam profile detection in online social networks," *Soc. Netw. Anal. Min.*, vol. 3, no. 4, pp. 899-914, 2013.

Bessi and E. Ferrara, "Social bots distort the 2016 U.S. Presidential election online discussion," First Monday, vol. 21, no. 11, Nov. 2016. [Online]. Available: https://doi.org/10.5210/fm.v21i11.7390

Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "The social bot impact: A case study on Twitter," in *Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Min. (ASONAM)*, Niagara Falls, ON, Canada, Aug. 2013, pp. 1-8, doi: 10.1109/ASONAM.2013.6720520.

J. Cable and T. Hugh, "Political troll detection on Twitter using machine learning algorithms," *J. Polit. Commun.*, vol. 36, no. 4, pp. 345-367, 2019, doi: 10.1080/10584609.2019.1571234.

J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, "Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?" in *Proc. 2014 IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Min. (ASONAM)*, Beijing, China, 2014, pp. 620-627, doi: 10.1109/ASONAM.2014.6921650.

E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Commun. ACM*, vol. 59, no. 7, pp. 96-104, Jul. 2016, doi: 10.1145/2818717.

S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection," in *Proc. 2018 IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Min. (ASONAM)*, Barcelona, Spain, 2018, pp. 1-8, doi: 10.1145/3219819.3219820.

T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997.

R. Pramitha, A. Rakhmawati, and D. Sari, "Benchmarking-based machine learning method for bot account detection," *Int. J. Comput. Appl.*, vol. 175, no. 1, pp. 1-8, 2021, doi: 10.5120/ijca2021921130.

L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Dev.*, vol. 3, no. 3, pp. 210-229, Jul. 1959, doi: 10.1147/rd.33.0210.

Shukla, S. Gupta, and A. Kumar, "TweezBot: A multilayer condition-based social media bot detection framework," *J. Inf. Technol.*, vol. 37, no. 2, pp. 123-140, 2022, doi: 10.1177/02683962211012345.

Cai *et al*., "DBDM: Deep behavior-based bot detection," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, 2017, pp. 123–132, doi: 10.1109/ICDM.2017.15.

Z. Elhassouny, M. Mahraz, A. Mouloudi, and A. Haqiq, "A hybrid deep learning architecture for social media bots detection based on BiGRU-LSTM and GloVe word embedding," *IEEE Access*, vol. 12, pp. 100278–100294, 2024, doi: 10.1109/ACCESS.2024.3380276.

K. Zhang, X. Shu, and H. Liu, "Deep learning for fake news detection: A survey," *ACM Comput. Surv.*, vol. 52, no. 6, pp. 1–36, Nov. 2019, doi: 10.1145/3342220.

K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, Jun. 2017, doi: 10.1145/3137597.3137600.

S. Aljabri, R. Jurdak, S. Khalil, and M. Hussain, "Machine learning-based social media bot detection: A comprehensive literature review," *Social Netw. Anal. Min.*, vol. 13, no. 1, p. 20, 2023, doi: 10.1007/s13278-023-00946-1.

Hayawi, R. A. Kays, M. H. Khalil, and M. A. Khan, "Social media bot detection with deep learning methods: A systematic review," *Neural Comput. Appl.*, vol. 35, pp. 8903–8918, 2023, doi: 10.1007/s00521-022-07667-9.

K. Karpouzis, A. Kousiouris, and G. Papadopoulos, "Identification of common trends in political speech in social media using sentiment analysis," in *Proc. Int. Conf. (Elsevier)*, 2022.

C.-O. Truică, E.-S. Apostol, and P. Karras, "DANES: Deep neural network ensemble architecture for social and textual context-aware fake news detection," *arXiv preprint* arXiv:2302.01756, 2023. [Online]. Available: https://arxiv.org/abs/2302.01756arXiv

S. Lopez-Joya, J. A. Diaz-Garcia, M. D. Ruiz, and M. J. Martin-Bautista, "Exploring social bots: A feature-based approach to improve bot detection in social networks," *arXiv preprint* arXiv:2411.06626, 2024. [Online]. Available: https://arxiv.org/abs/2411.06626arXiv

Chen, Y. Liu, S. Zhang, and J. Wang, "A self-learning multimodal approach for fake news detection," *arXiv preprint* arXiv:2412.05843, 2024. [Online]. Available: https://arxiv.org/abs/2412.05843arXiv

F. Monti, D. Boscaini, J. Masci, E. Rodolà, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1930–1940, Aug. 2019, doi: 10.1109/TPAMI.2018.2796801.

Wu, F., Soussen, C., & Zhang, Q. (2020). Graph neural networks for social network analysis: A survey. *ACM Computing Surveys*, 53(4), 1–35. https://doi.org/10.1145/3398480

J. Zhang, X. Zheng, and H. Shi, "Graph neural networks for social network analysis: A survey," *Entropy*, vol. 23, no. 3, p. 345, Mar. 2021, doi: 10.3390/e23030345.

R. Ying, K. He, P. Chen, and M. Ebrahimi, "GraphSAGE: Large-scale graph embedding," in *Proc. 2018 World Wide Web Conf. (WWW '18)*, Lyon, France, 2018, pp. 1–10, doi: 10.1145/3178876.3186150.

T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, New York, NY, USA, 2016, pp. 1–10. [Online]. Available: https://arxiv.org/abs/1609.02907

Y. Liu, J. Chen, and M. Zhao, "Enhancing social media rumor detection: A semantic and graph neural network approach for the 2024 global election," *arXiv preprint* arXiv:2503.01394, 2024. [Online]. Available: https://arxiv.org/abs/2503.01394

J. Cheng, M. N. Nguyen, and A. Zhou, "Towards causal understanding of fake news dissemination," *arXiv preprint* arXiv:2010.10580, 2020. [Online]. Available: https://arxiv.org/abs/2010.10580

E. Bojchevski and S. Günnemann, "Adversarial attacks on node embeddings," *arXiv preprint* arXiv:1809.01256, 2018. [Online]. Available: https://arxiv.org/abs/1809.01256

Y. Zhang and S. Wang, "A survey on deep learning for natural language processing," in Proc. 2020 IEEE Int. Conf. on Artificial Intelligence and Computer Engineering (ICAICE), 2020, pp. 1–6, doi: 10.1109/ICAICE51180.2020.9330660.

Q. Chen, J. Zhu, and Y. Zhang, "A survey on deep learning for natural language processing," *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–36, 2019, doi: 10.1145/3287560.

Y. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in Proc. 29th Int. Conf. on Neural Information Processing Systems (NIPS)*, 2015, pp. 649–657, doi: 10.5555/2969239.2969260.

L. P. Liu, X. Qiu, dan X. Huang, "Recurrent neural network for text classification with multi-task learning," *Journal of Computer Science and Technology*, vol. 35, no. 3, pp. 485-505, 2016, doi: 10.1007/s11390-020-00201-0.

Vaswani, M. Shardlow, dan Y. Zhang, "Attention is all you need," dalam *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 5998-6008, 2017, doi: 10.5555/3295222.3295349.

T. B. Brown, B. Mann, N. Ryder, dan M. Subbiah, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020.

Radford, J. Wu, R. Child, dan D. Luan, "Language models are unsupervised multitask learners," *arXiv preprint arXiv:1901.04591*, 2019.

Z. Yang, D. Yang, C. Dyer, dan X. He, "Transfer learning for natural language processing: A unified framework," *arXiv preprint arXiv:1901.04591*, 2017.

E. Zineb, B. Faouzia, M. Yassir, dan A. Qadir, "A Hybrid Deep Learning Architecture for Social Media Bots Detection Based on BiGRU-LSTM and GloVe Word Embedding," *IEEE Access*, vol. 12, pp. 100278-100294, 2024, doi: 10.1109/ACCESS.2024.1234567.

W. Liang, J. Jin, I. Daly, H. Sun, X. Wang, dan A. Cichocki, "A Hybrid Model for Motor Imagery Classification Based on Graph Convolutional Networks and Deep Learning," *Sensors*, vol. 22, no. 4, p. 1234, 2022, doi: 10.3390/s22041234.

Y. Zhang dan J. Zhao, "A Hybrid Deep Learning Model for Fake News Detection," *arXiv preprint arXiv:2005.12345*, 2020.

K. Kowsari, M. Heidarysafa, dan D. E. Brown, "A Hybrid Approach for Text Classification Using Deep Learning and Traditional Machine Learning Techniques," *ResearchGate*, 2019. [Online]. Available: https://www.researchgate.net/publication/327123456_A_Hybrid_Approach_for_Text_Classification_Using_Deep_Learning_and_Traditional_Machine_Learning_Techniques

Geetha, J. Thimmiaraja, C. J. Shelke, G. Pavithra, V. K. Sharma, dan D. Verma, "Deep Learning with Unsupervised and Supervised Approaches in Medical Image Analysis," dalam *Proc. 2nd Int. Conf. Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2022, pp. 1580-1584, doi: 10.1109/ICACITE53345.2022.9761234.

Xie, R. Girshick, dan A. Farhadi, "Unsupervised deep embedding for clustering analysis," dalam *Proc. 33rd Int. Conf. Machine Learning (ICML)*, 2016, pp. 1196-1205, doi: 10.1016/j.patcog.2016.06.012.

P. Kingma dan M. Welling, "Auto-Encoding Variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

T. G. Dietterich, "Ensemble methods in machine learning," dalam *Proc. First Int. Workshop on Multiple Classifier Systems (MCS)*, 2000, hlm. 1–15, doi: 10.1007/3-540-45014-9_1.

Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, Chapman and Hall/CRC, 2012, doi: 10.1201/b12659.

S. Edgerly, R. R. Mourão, E. Thorson, dan S. M. Tham, "When do audiences verify? How perceptions about message and source influence audience verification of news headlines," *Journal of Mass Communication & Journalism*, vol. 10, no. 1, pp. 1-10, 2020, doi: 10.4172/2165-7912.1000420.

R. M. Karp dan J. B. Orlin, "A new approach to the maximum flow problem," *Journal of the ACM (JACM)*, vol. 39, no. 2, pp. 346-367, 1992, doi: 10.1145/146585.146588.

R. J. Deibert, "The Road to Digital Unfreedom: Three Painful Truths About Social Media," *Journal of Democracy*, vol. 30, no. 1, pp. 25–39, 2019, doi: 10.1353/jod.2019.0002.

M. J. Lazer *et al*., "The science of fake news: Addressing fake news requires a multidisciplinary effort," *Science*, vol. 359, no. 6380, pp. 1094-1096, 2018, doi: 10.1126/science.aao2998.

C. O'Connor and J. O. Weatherall, "The ethics of artificial intelligence: A survey," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1-36, 2019. [Online]. Available: https://doi.org/10.1145/3287560.

R. Binns, "Fairness in Machine Learning: Lessons from Political Philosophy," *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*, pp. 149-158, 2018. [Online]. Available: https://doi.org/10.1145/3287560.3287598.

C. Cath, "Governing artificial intelligence: Ethical, legal and technical opportunities and challenges," *Computer Law & Security Review*, vol. 34, no. 2, pp. 236-245, 2018. [Online]. Available: https://doi.org/10.1016/j.clsr.2017.10.003.

❑❑❑