AI-enabled Bot and Social Media: A Survey of Tools, Techniques, and Platforms for the Arms Race

Flavio Lombardi[1] Maurantonio Caprolu[2]
Roberto Di Pietro[2]

[1]Istituto per le Applicazioni del Calcolo, Consiglio Nazionale delle Ricerche (IAC-CNR) Rome, Italy
[2]Division of Information and Computing Technology
College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

Abstract

AI-enabled bots are quietly affecting the lives of billions of people. As an example, Covid-19 pandemics has shown that information can be a deadly weapon: the uncertain and sometimes outright false information on vaccines has generated in the general public some diffidence that has slowed down vaccination causing loss of lives.

The interplay between the trustworthiness of information, the propagation of information and, especially, fake news, has drawn the attention of the research community, with a few research contributions shedding light on the topic. Such research has shown that a vast part of misinformation has been developed through bots that can actively spread misinformation and at the same time slow down (if not stop) the diffusion of legitimate/genuine information. Bot technologies and implementations have flourished and are constantly evolving in order to avoid detection and to increase their reach and impact. At the same time OSNs have put in place increasingly sophisticated countermeasures to limit bots.

This chapter reviews the latest AI techniques supporting bot, as well as its dual, the most relevant AI inspired techniques for bot detection. The systematization provided in this contribution has the objective to categorize the approaches that support AI-enabled bot, to shed light on the most promising techniques to detect them, as well as to provide some future research directions.

This is a personal copy of the authors. Not for redistribution. The final version of the paper is available as part of the book entitled *"Mixed Methods Perspectives on Communication and Social Media Research"*, edited by Reynaldo Gacho Segumpan and Joanna Soraya Abu Zahari, ISBN 9781032209128, Chapter 15. To cite this chapter, download the bibtex here.

1 Introduction to Bots

The impact of AI-enabled bot is a subject of discussion in academic venues, while the general public is not much aware of it. However, the topic is affecting the lives of hundreds of millions, if not billions, of people. The best example of such an impact is related to the chronicles of the last one year and a half; Covid-19 has triggered an acceleration to many dimensions of our everyday life: digitization, mass safety measures, international coordination (to an extent), but it has also shown the main weaknesses of our interconnectedness, primarily related to Online Social Networks (OSNs). In particular, it has highlighted how the current information ecosystem promotes the rapid dissemination of news to the detriment of their trustworthiness and reliability. As a result, information is frequently used as a weapon to alter reality, spreading fake news with, often, severe effects. As an example, the unverified, contradictory, and sometimes outright false information on the Covid-19 pandemic has generated a certain degree of diffidence in the general public, slowing down the vaccination campaign. As a result, the pandemic has continued to ramp up worldwide, diminishing the positive effects of vaccine immunization, and increasing the life toll.

The spread of fake news on OSNs has been under investigation for several years by the research community. Nevertheless, the nefarious effects of misinformation on the Covid-19 vaccination campaign have rekindled interest in this topic. A few research studies shed light on this malicious practice, revealing one of the widespread digital weapons used to spread disinformation: socialbots. bots are certainly not a new technology. Defined as software designed to automate a set of tasks, we can cite many example of bots, such as automated web crawlers and HTTP load/request creators. Similarly, a socialbot is a computer algorithm designed to run automated tasks on specific social contexts, such as OSNs, and directly interact with humans.

Initially, socialbots were designed to perform repetitive tasks faster and more efficiently than humans, such as automated content aggregators. Subsequently, socialbots have evolved to emulate and sometimes alter human behavior, pushed by the rapid advancement of AI technologies. Depending on their scope and technological level, socialbots can have different abilities and perform different tasks, ranging from very simple to highly complex activities. For example, the new generation of socialbots can autonomously talk with a human via online chat and spoken language, manage an OSN account, and trade online on financial platforms. On the one hand, this technology enables the development of new and exciting software capabilities, such as the full automation of corporate customer care services. On the other hand, socialbots can sometimes be harmful. As an example, bots can actively spread misinformation online, slowing down (if not even stopping) the diffusion of truthful information. In addition, bots are ever-evolving to deceive the increasingly sophisticated countermeasures that OSNs are deploying for detecting and stopping automated malicious activities.

To fully understand the capabilities of the new generation of socialbots and what is needed to stop their malicious activities, it is essential to know the tech-

nologies that are pushing their evolution. With this goal in mind, in this chapter, we provide a review of the latest AI technologies supporting bots. For each considered technology, we provide a concise overview, explaining the advances it is providing. Furthermore, we discuss the malicious activities of socialbots documented in the literature and review the state-of-the-art of existing detection methodologies. In addition, we also highlight some related research directions.
Roadmap. The rest of this chapter is organized as follows. In Section 2 we analyze the AI-based technological pillars supporting socialbots, present the state-of-the-art of existing socialbots, and we list the most important toolkits, frameworks, and SDKs publicly available for developing socialbots. In Section 3 we briefly summarize the latest research addressing malicious bot, while in Section 4 we review the state-of-the-art of bot detection. Finally, Section 5, draws some final remarks.

2 AI and Socialbot

As discussed in the previous section, a socialbot is nothing more than a bot that operates on specific "social" contexts, such as OSNs. As such, a socialbot is not necessarily advanced software able to run complex activities thanks to AI-based modules. We can mention several examples of non-AI-based socialbots, such as Twitter bots that automatically post the latest news related to specific topics. In fact, automating basic tasks in OSNs, such as following users, re-posting specific content, and liking/disliking other posts, is doable without involving AI-based technologies. However, emerging use cases require more than basic tasks automation. Recent advancements in AI technologies have enabled the development of a new generation of socialbots that can emulate human behavior. Companies widely use this technology to manage their social media accounts automatically, answer customers' questions, and automate customer care services. Nevertheless, AI-based socialbots are also used for illegal and unethical purposes, such as spreading fake news and attempting to manipulate stock markets. Among the different types of socialbots, we can identify the following categories (Boshmaf, Muslukhov, Beznosov, & Ripeanu, 2013):

- *Self-declared bots:* Legitimate and useful bots, e.g., Twitter bots that post latest news related to specific topics.
- *Spambots:* bots that massively distribute unrequested messages, mainly advertisement, to unwilling users.
- *Human-like bots:* bots designed to behave as a human being.

In this section, we discuss AI-based socialbots by analyzing their technological pillars and their recent advancements. First, we discuss the AI technologies that have been fundamental for developing the next generation of socialbots. Then, we present the state-of-the-art socialbots and list the tools publicly available for socialbot development.

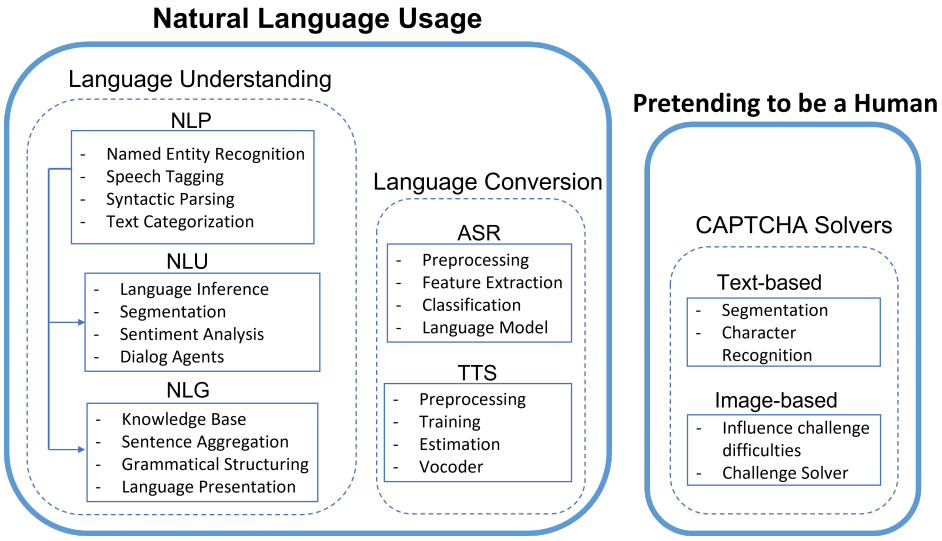

Figure 1: AI techniques supporting socialbots

2.1 AI Techniques Supporting Socialbots

This section discusses the most prominent AI technologies, depicted in Figure 1, that have enabled the development of a new generation of socialbots designed to emulate human behavior. First, NLP and its subfields played a crucial role in socialbots evolution, making them understand natural language. Next, natural language conversion methodologies, such as Automatic Speech Recognition (ASR) and Text-To-Speech (TTS), completed the puzzle giving socialbots the ability to interact with humans. Finally, advances in computer vision have enabled the development of CAPTCHA solvers, giving socialbots the ability to pass automated tests designed to tell computers and humans apart.

2.1.1 Natural language Processing

The goal of making computers understand human language has driven research in linguistics and computer science since the '50s of last century. In those years, the automatic interpretation and generation of natural language began to attract the attention of researchers, laying the foundations for a wide area of study called Natural Language Processing (NLP). Initially, the main goal was to create a computer capable of extracting and interpreting the information contained in digital documents. Subsequently, the focus shifted to the direct interaction between computers and human beings through natural language. With these aims, NLP combines different techniques from several disciplines of linguistic and computer science to make computers understand natural language, in both its written and verbal forms. From a high-level perspective, NLP takes unstruc-

tured data in input and converts them into structured formats. This concept has been implemented using multiple techniques over the years. However, three distinct methodologies can be identified to represent the evolution of NLP from its beginnings to its present form. Initially, the prevailing approach, called symbolic NLP, was based on the definition of a set of handwritten rules. Then, the software applied this complex set of rules to the input data, performing basic NLP tasks. Subsequently, around the 90s, there was an evolution of NLP methodologies due to the advent of the first artificial intelligence technologies. The symbolic NLP was replaced by the statistical NLP, driven by the development of novel ML techniques and the increasing number of digitized documents available for training ML algorithms. Finally, modern NLP systems heavily rely on deep learning methods, achieving state-of-the-art results in what is considered the last revolution in this field, called Neural NLP.
NLP includes several tasks, methodologies, and subfields. The most important and relevant for our discussion on socialbots are Natural Language Understanding (NLU) and Natural Language Generation (NGU).

NLU provides the multiple text analyses necessary for computers to understand the intended meaning of a sentence/conversation. NLP methodologies analyze both the syntax and the semantic of the text received as input. In addition, NLU also provides a relevant ontology, i.e., a data structure which specifies the relationships between words and phrases. NLU is commonly used for sentiment analysis, a methodology for clustering positive and negative comments on social media and customer feedback.
NLG enables computers to generate text. From a high-level perspective, NLG uses AI techniques to produce a human-readable text response based on the structured data received as input. The first AI-based NLG techniques relied on ML algorithms, such as Markov chains, to predict the next word in a sentence. Subsequently, NLG methodologies moved to deep learning techniques, such as Recurrent Neural Networks (RNN) and Long short-term Memory (LSTM), achieving better results.

2.1.2 Natural Language Conversion

Thanks to NLP, computers can autonomously understand and use natural language. However, NLP alone is not enough for allowing software to perform a complete interaction with humans. For this purpose, at least two other pieces are needed to complete the puzzle: ASR and TTS.

ASR refers to a set of methodologies that allow recognizing a speech and converting it into text format starting from an audio track. Conversely, TTS refers to a set of tools that enable computers to speak with human-like voices starting from text. A typical ASR architecture consists of three modules. The first module is dedicated to pre-processing the audio track in order to remove any noise registered alongside the audio. Then, after receiving the clean audio signal, the second module extracts the features that will be used for classification. The principal methodologies used for this task are Melfrequency Cepstral Coefficients (MFCCs), Linear Predictive Coding (LPC), and Discrete Wavelet

Transform (DWT). Finally, the classification module uses the extracted features to predict the text corresponding to the input speech signal. Among the most used classification models for speech recognition, we can cite HMM and Gaussian mixture models (GMM), Support Vector Machines (SVM), and Artificial Neural Networks (ANN). A fourth optional module, called language module, can significantly improve efficiency by considering various types of rules and semantics of a language (Malik, Malik, Mehmood, & Makhdoom, 2021).

A common architecture of a TTS system consists of two main components: text processing and methods of speech generation. The input of such a system is plain text in the form of a chain of words. This input may include words available in the dictionary and several other components, such as punctuation symbols, numbers, abbreviations, and possibly others. For this reason, the input text needs to be pre-processed. This step includes the normalization of any non-standard word and the prediction of the appropriate intonation and generation of a phonetic sequence for each considered word. The second module, called speech generation, includes a set of tools to convert phonetic sequences into speech waveform by using speech synthesis techniques (Prahallad, 2010). This technology has evolved similarly to NLP. Initially, the first systems were based on the concatenation of pre-recorded sounds, i.e., concatenation synthesis. Then, artificial intelligence has brought a revolution also in this field with the advent of the so-called Statistical Parametric Synthesis (SPS). An SPS system generally works in two distinct steps. First, a training step is performed to characterize a large set of audio samples, i.e., the dataset. During this phase, several parameters are extracted, such as the frequency and the duration. Then, a statistical model is used to estimate those parameters that are finally reconverted to speech waveform by using rebuild approaches, i.e., vocoders. SPS introduces significant advantages and improves the quality of the synthesized speech over concatenation synthesis. However, a novel generation of vocoders based on deep learning techniques sensibly overcomes every previous methodology. State-of-the-art vocoder proposals (Jiao et al., 2021) can synthesize a wide range of voices, styles, and languages leveraging neural networks. Parallel WaveNet (Oord et al., 2018), for example, can convert a sequence of input noise into audio waveforms leveraging parallel computing. As a result, it can synthesize samples very efficiently by fully exploiting the computational power of modern deep learning hardware.

2.1.3 AI-based Approaches for Solving CAPTCHAs

Cybercriminals increasingly use malicious bots to automate different types of cyberattacks such as credential stuffing, illicit registration of multiple free accounts, malicious crawling, and possibly others against online services. One of the most effective countermeasures, called Completely Automated Public Turing tests to tell Computers and Humans Apart (CAPTCHA), has been used since 1996 to distinguish between legitimate users and bots (Guerar, Verderame, Migliardi, Palmieri, & Merlo, 2021). The first proposals were based on the inability of Optical Character Recognition (OCR) software to recognize a

distorted text. Subsequently, novel schemes have been proposed following the tremendous progress of AI-based techniques in computer vision and similar fields that makes solving CAPTCHAs easier for bots. Ideally, such problems, called AI-hard, cannot be solved by machines but can be easily solved by simple human interactions (Ali, Caprolu, & Di Pietro, 2020). However, increasingly accurate AI-based models can be used to circumvent this countermeasure. In fact, several proposals can be found in the literature to solve image-based CAPTCHAs, still the most widely used in many commercial web services despite the availability of more robust schemes (Bursztein, Martin, & Mitchell, 2011). The most prevalent approach consists of two different phases. First, CAPTCHA's image is segmented into single characters. Then, those characters are recognized individually using ML algorithms such as Support Vector Machine (SVM) and Convolutional Neural Networks (CNN). This approach works well for simple CAPTCHAs, without complex security features such as occluding lines and distorted characters. However, for more advanced schemes, (Wang, Wei, Zhang, Liu, & Wang, 2021) proposed a captcha transformation model based on cycle-consistent Generative Adversarial Network (GAN) that reduces the segmentation difficulties even in the presence of complex security features. The proposed model adopts two mirror-symmetrical GANs that, after an optimization phase, can transform the original image into a simplified one by removing background noise. Then, the segmentation and recognition steps are performed on the simplified image by using traditional ML techniques. Another commonly used approach is based on dictionaries. In (Chougule, Tupsamudre, & Lodha, 2020), the authors proposed Revelio, an efficient CAPTCHA solver that relies only on image processing techniques. The training phase requires at most 100 labeled images to create the dictionary. Then, the solving phase removes the background noise, segments the image, and recognizes the characters using either the image's hash or its histogram of oriented gradient.

Given the evident weaknesses of text-based CAPTCHAs, google released a more sophisticated scheme, called reCaptcha, that includes image-based challenges. However, several mechanisms for breaking even this kind of puzzle appeared quickly. In (Sivakorn, Polakis, & Keromytis, 2016), the authors presented a complete analysis of reCaptcha and proposed a novel scheme to solve it automatically. Their scheme is based on two main steps. First, they influence the process that determines the level of difficulties of the puzzle, dynamically computed by the system depending on several aspects of the client's browser. Then, they analyze the challenge and use several techniques to solve it, depending on its type. The primary solving mechanism is based on image classification techniques using Deep convolutional Neural Networks (DNN). Unfortunately, even after a major security update of reCaptcha by Google, the system is still vulnerable to several attacks enabled by recent advances in Object Detection algorithms, as demonstrated by (Hossen et al., 2020).

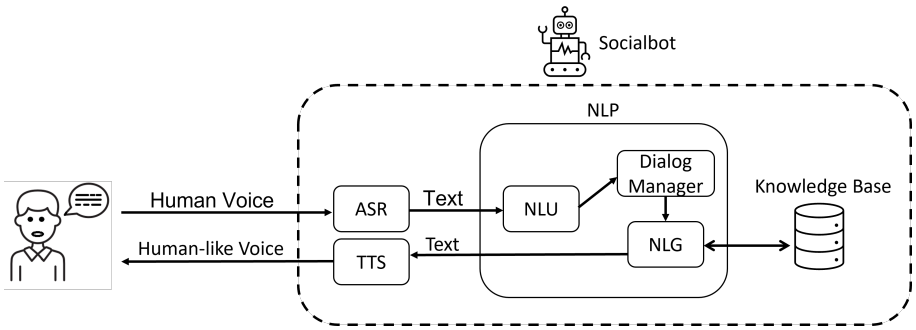

Figure 2: Common socialbot architecture

2.2 AI-based Socialbots

The last decade has seen an increasing evolution of socialbots driven by the technological advances discussed in Section 2.1. The main goal of the new generation of socialbot is to communicate with humans through natural language, in both its written and spoken forms. Although many existing socialbots are proprietary software, we can also find several proposals in the literature. Some of them can conduct an online chat conversation with humans, while others can even communicate in the spoken language. To the best of our knowledge, there are no malicious socialbot design proposals in the literature. However, the available chatbot designs could be easily used by cybercriminals to develop malicious socialbots.

In (Xu, Liu, Guo, Sinha, & Akkiraju, 2017), the authors proposed a novel chatbot for customer service on OSNs. Their proposal suggests that the interaction between a chatbot and a customer can be simplified to the problem of mapping a sequence of words, which represent the request, to another sequence of words, which represent the answer. To learn the mapping function, they applied deep learning techniques, such as Long Short-Term Memory (LSTM) networks. In (Ranoliya, Raghuwanshi, & Singh, 2017), the authors leveraged Artificial Intelligence Markup Language (AIML) and Latent Semantic Analysis (LSA) for developing a chatbot tailored for university-related FAQs. A research team from the University of Montreal proposed MILABOT (Serban et al., 2017), a deep reinforcement learning chatbot. They developed a large-scale ensemble-based dialogue system framework that leverages several ML methodologies, such as deep learning and reinforcement learning. They proposed novel deep learning models for NLG, including recurrent neural networks, sequence-to-sequence models, and latent variable models.

Spoken conversational agents are becoming more and more pervasive in modern society (Caprolu, Sciancalepore, & Di Pietro, 2021). We can cite many examples of this type of socialbot, such as personal assistants, e.g., Amazon Alexa and Apple Siri. Although the rapid evolution of this technology, building "intelligent" conversational agents remains an open research problem. State-

of-the-art AI-based assistants focus on short interactions typically related to a limited set of available contexts. In (Fang et al., 2018), the authors proposed Sounding Board, a chatbot designed to hold coherent and engaging long-term conversations. The architecture consists of an NLU module that analyzes the user speech, a dialog manager executes the dialog policy and decides the next dialog state, and an NLG module that builds the response. Further improvements come from (Yu et al., 2019), where the authors proposed Gunrock, a chatbot able to generate more stable conversations by encouraging and understanding more human inputs. The system architecture is based on the Amazon Conversational Bot Toolkit (CoBot), an event-driven framework that provides ASR and NLP support. In addition, Gunrock proposed a novel NLU module to analyze users' utterances. Then, a dialog manager uses NLU techniques to select the appropriate topic and defines the dialog flow. Finally, an NLG module generates the appropriate response leveraging multiple knowledge bases, and a TTS module returns the synthesized response to the user.

2.3 Socialbot Developing Platforms

In this section, we list and briefly describe the most important publicly available toolkits, frameworks, and SDKs for developing AI-based socialbots. Given the rapid spread of emerging use cases that require socialbots, both in the commercial and academic fields, it is possible to find many development solutions to create even highly sophisticated socialbots. Several commercial platforms, such as Microsoft Azure Bot Service[1] and Amazon Lex[2], enable developers to design and develop enterprise-grade conversational AI agents quickly and easily. Among the open-source solutions, we can cite the Microsoft Bot Framework SDK[3]. With this SDK, developers can build advanced conversational agents leveraging a large set of NLU, NLG, and QnA maker modules. The Microsoft Bot Framework supports multiple programming language, such as C#, JS, Python and Java. In addition, it provides a visual authoring canvas for developing socialbots without dealing with underlying layers. In 2018, Amazon introduced CoBot(Khatri et al., 2018), a conversational bot toolkit that hides implementation and infrastructure layers to allow developers to concentrate more on science challenges. CoBot sensibly cut the time to develop socialbots by providing interfaces to several Amazon services like cloud computing and NLP modules. Another open-source methodology for creating AI-based chatbots is Artificial Intelligence Markup Language (AIML), an XML schema for specifying heuristic conversation rules. An AIML file can be run by using an interpreter, such as Program AB[4]. In addition, the kernel of the Artificial Linguistic Internet Computer Entity (A.L.I.C.E.)(Wallace, 2009), the first chatbot developed in

[1] https://azure.microsoft.com/en-us/services/bot-services/
[2] https://aws.amazon.com/lex/
[3] https://github.com/microsoft/botframework-sdk
[4] https://code.google.com/archive/p/program-ab/

AIML, is open-source and freely available online[5]. This kernel is a set of AIML templates tan can be used as a platform to develop chatbots. Another interesting framework for socialbot developers is Text-to-Speech for all [6], a library for advanced Text-to-Speech generation distributed by Mozilla.

3 Malicious Bots

Sophisticated bots, i.e., automated software that can mimic human behavior and deceive conventional security measures, increased 18% in 2019, and now account for 45% of the overall malicious bot web traffic (Radware, 2020). AI-based socialbots can be used to perform several malicious activities on the virtual domain, ranging from misinformation to stock market manipulation. In (Tardelli, Avvenuti, Tesconi, & Cresci, 2021), the authors investigated financial disinformation on OSNs. They have shown that socialbots often artificially inflated the popularity of the most viral stocks on Twitter. Then, they proposed two different AI-based methodologies for detecting financial disinformation campaign on Twitter, via classification and regression. Another malicious activity, even if technically not illegal, was discussed in (Somanath, 2021). The authors investigated how AI-based bots are used by scalpers to quickly empty out inventories by online sellers to then sell them for more than double the price.

The large amount of fake news appearing on OSNs in recent years has drawn the attention of the research community. In (Zhang & Ghorbani, 2020), the authors presented a survey of the findings to date relating to misinformation in OSNs. They characterized the negative impact of the fake news online spreading, and the state-of-the-art in detection methods. Finally, they discussed the different approaches used for detection, ranging from fact-checking websites to scientific methods based on academic research. In (Shao et al., 2018), the authors showed that socialbots play a disproportionate role in amplifying and spreading articles from low-credibility sources. Furthermore, they highlighted how humans are vulnerable to this manipulation, re-sharing the content posted by bots. Given that socialbots heavily support successful low-credibility sources, the authors showed that curbing socialbots may be an effective strategy for mitigating the spread of online misinformation. To help understand the dynamics of fake news dissemination in OSNs, (Guarino, Trino, Chessa, & Riotta, 2020) presents the DisInfoNet Toolbox. DisInfoNet combines text mining and classification with graph analysis and visualization to offer a comprehensive and user-friendly suite for tracking the origin and the broadcasters of false information. DisInfoNet can be used to track relevant news stories and reconstruct their prevalence over time and space; to detect central debating communities and capture their distinctive polarization/narrative; and finally to identify influencers both globally and in specific "disinformation networks".

The malicious activities discussed above are only a small subset of the wide range that can be performed by using socialbots. Many other malicious

[5] https://code.google.com/archive/p/aiml-en-us-foundation-alice/
[6] https://github.com/mozilla/TTS

tasks can be automated, such as phishing and credential stuffing, credit cards testing (Somanath, 2021), and cryptojacking (Caprolu, Raponi, Oligeri, & Di Pietro, 2021), just to cite a few. In addition socialbots, thanks to AI-based technologies, are showing an ever-increasing ability to deceive existing security measures. For these reasons, the prompt detection of socialbots has become of paramount importance for modern cybersecurity systems.

4 Bot detection

Similarly to what above discussed, the approaches and techniques devoted to detecting bots have become increasingly sophisticated over time. In the following we describe some of the most relevant ones in chronological order, together with their underlying technologies.

The BotPrize (Hingston, 2009) competition environment and testing protocol has often been used as a method in research efforts for assessing believability. One of such efforts is (Asensio et al., 2014) where some bots were tested using a cognitive approach taking into account human psychology with an approach aimed at imitating human behaviour at the nervous system level.

BotOrNot (now Botometer) (Davis, Varol, Ferrara, Flammini, & Menczer, 2016) is a tool created in 2016 and leveraging thousands of features to evaluate the similarity of a Twitter account to the known characteristics of socialbots. The BotOrNot classifier uses Random Forest, an ensemble supervised learning approach.

Cresci (Cresci, Di Pietro, Petrocchi, Spognardi, & Tesconi, 2017) in 2017 compared social spam bots with previously known spambots and evaluated Twitter countermeasures effectiveness also benchmarking state-of-the-art techniques proposed in academic literature.

Botnet detection frameworks surveyed in (Alieyan, Almomani, Manasrah, & Kadhum, 2017) focus on DNS traffic analysis for detection. The outcome is that recent botnets do not rely on static IP addresses but resort to DNS to evade detection. To improve DNS based botnet detection, DNS-BD (DNS rule-based approach to botnet detection) was proposed in (Alieyan et al., 2021) where the approach involves analysing DNS traffic, and seeing if the behavior of some nodes matches given rules, such as coordination in requesting the same IP, and detecting abnormalites in the DNS based botnet system.

A deep neural network based on contextual LSTM architecture that exploits both content and metadata to detect bots at the tweet level was proposed by (Kudugunta & Ferrara, 2018). Synthetic minority oversampling was used to generate a large labeled dataset, suitable for training deep nets from a small amount of labeled data allowing to achieve high classification accuracy (AUC>96%) from just a single tweet in the dataset provided by (Cresci et al., 2017).

A framework for detecting advanced web bots was presented in (Iliou et al., 2019) using a list of features such as session time, browsing speed, and GET/POST/HEAD requests, that are run through a supervised classification

method. Given that the problem of detecting web bots is similar to detecting socialbots, some of the proposed countermeasures can be used pervasively.

Botscan (Wirth, Menchen-Trevino, & Moore, 2019) is a tool to track bots at the conversation-level and in real-time. Botscan aims to assist researchers, journalists, and the general public to establish if a deliberation in online discourse is driven by authentic or fake communication by measuring bot activity at the conversation level in real-time, supported by the BotOMeter algorithms (Sayyadiharikandeh, Varol, Yang, Flammini, & Menczer, 2020). Nevertheless, as shown in (Rauchfleisch & Kaiser, 2020) Botometer was imprecise over time for the new bots and a dataset in German language. As such Botometer reliability for social science research is at stake.

A bot detection model for discriminating between human user's and bot web access behavior is discussed in (TANAKA et al., 2020). The approach leverages logistic regression and LightGBM (Lebeuf, Zagalsky, Foucault, & Storey, 2019).

The bot/spammer detection proposed in (Pham, Nguyen, Vo, & Yun, 2021) leverages network representation learning (NRL) to create Bot2Vec, to evaluate the features of relations between OSN users, not considering the features of users' profiles. The objective is to capture both the neighborhood and community-aware structure of each OSN user, based on the assumption that normal accounts and bots/machine accounts are frequently active in their own communities.

The detection approaches above discussed are a relevant subset of the large amount of efforts devoted to addressing the problem. In other words, the smarter the attacker, the smarter has to be the detection capability. The challenge ahead is to leverage advanced AI more effectively to fasten detection and to improve precision and recall, also to limit misuse of the detection system itself that would possibly affect legitimate users.

5 Conclusion

We have started this chapter starting from the observation that on-line social networks are influencing more and more people behaviour, typically framing the context for any given topic of discussion. In this scenario, the role of bot is a pre-dominant one. In particular, we have analyzed the evolution of socialbots, characterized their objectives and behaviour, and spot-lighted malicious socialbot and bot detection platforms and techniques.

If we have generalize what discussed in this chapter, it appears that bots have evolved from simple deception mechanisms to fully fledged AI-empowered software. At the same time bot detection approaches have matured and are now able to perform a much more elaborate and comprehensive analysis of the combined characteristics of collaborating/coalitions of bots. This never ending war is far from being over. It is quite the opposite: there is an urgent challenge for Industry and Academia that calls to invest time and resources in the field, with the objective to magnify the useful and legitimate usage of bot, while fighting the dark side of them. We believe that in this manuscript

we contribute to help shedding light on this challenging domain, and provide pointers to resources to better understand present best practices and future research trends for both sides of the challenge.

Acknowledgements

This contribution was partially supported by the TAILOR Project - EU HORIZON 2020 Research and Innovation Programme GA No 952215 (https://tailor-network.eu). This contribution was also partially supported by awards NPRP-S-11-0109-180242 from the QNRF-Qatar National Research Fund, a member of The Qatar Foundation, and NATO MYP G5828 project "SeaSec: DronNets for Maritime Border and Port Security". The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the cited sponsors.

We are especially grateful to our colleague Dr. Stefano Guarino for his suggestions.

References

Ali, I. M., Caprolu, M., & Di Pietro, R. (2020, August). Foundations, properties, and security applications of puzzles: A survey. *ACM Comput. Surv.*, *53*(4). Retrieved from https://doi.org/10.1145/3396374 doi: 10.1145/3396374

Alieyan, K., Almomani, A., Anbar, M., Alauthman, M., Abdullah, R., & Gupta, B. B. (2021). Dns rule-based schema to botnet detection. *Enterprise Information Systems*, *15*(4), 545-564. doi: 10.1080/17517575.2019.1644673

Alieyan, K., Almomani, A., Manasrah, A., & Kadhum, M. M. (2017, July). A survey of botnet detection based on dns. *Neural Comput. Appl.*, *28*(7), 1541–1558. Retrieved from https://doi.org/10.1007/s00521-015-2128-0 doi: 10.1007/s00521-015-2128-0

Asensio, J. M. L., Peralta, J., Arrabales, R., Bedia, M. G., Cortez, P., & Peña, A. L. (2014). Artificial intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters. *Expert Systems with Applications*, *41*(16), 7281–7290.

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2013). Design and analysis of a social botnet. *Computer Networks*, *57*(2), 556-578. Retrieved from https://www.sciencedirect.com/science/article/pii/S1389128612002150 (Botnet Activity: Analysis, Detection and Shutdown) doi: https://doi.org/10.1016/j.comnet.2012.06.006

Bursztein, E., Martin, M., & Mitchell, J. (2011). Text-based captcha strengths and weaknesses. In *Proceedings of the 18th acm conference on computer and communications security* (p. 125–138). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2046707.2046724 doi: 10.1145/2046707.2046724

Caprolu, M., Raponi, S., Oligeri, G., & Di Pietro, R. (2021). Cryptomining makes noise: Detecting cryptojacking via machine learning. *Computer Communications*, *171*, 126-139. Retrieved from https://www.sciencedirect.com/science/article/pii/S0140366421000797 doi: https://doi.org/10.1016/j.comcom.2021.02.016

Caprolu, M., Sciancalepore, S., & Di Pietro, R. (2021). Short-range audio channels security: Survey of mechanisms, applications, and research challenges. *IEEE Communications Surveys Tutorials*, *23*(1), 311-340. doi: 10.1109/COMST.2020.2969030

Chougule, A., Tupsamudre, H., & Lodha, S. (2020). Revelio: A lightweight captcha solver using a dictionary based approach. In S. Kanhere, V. T. Patil, S. Sural, & M. S. Gaur (Eds.), *Information systems security* (pp. 97–116). Cham: Springer International Publishing.

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm-shift of social spambots. *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*. Retrieved from http://dx.doi.org/10.1145/3041021.3055135 doi: 10.1145/3041021.3055135

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botornot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web* (p. 273–274). Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee. Retrieved from https://doi.org/10.1145/2872518.2889302 doi: 10.1145/2872518.2889302

Fang, H., Cheng, H., Sap, M., Clark, E., Holtzman, A., Choi, Y., ... Ostendorf, M. (2018). *Sounding board: A user-centric and content-driven social chatbot.*

Guarino, S., Trino, N., Chessa, A., & Riotta, G. (2020). Beyond fact-checking: Network analysis tools for monitoring disinformation in social media. In H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, & L. M. Rocha (Eds.), *Complex networks and their applications viii* (pp. 436–447). Cham: Springer International Publishing.

Guerar, M., Verderame, L., Migliardi, M., Palmieri, F., & Merlo, A. (2021, October). Gotta captcha 'em all: A survey of 20 years of the human-or-computer dilemma. *ACM Comput. Surv.*, *54*(9). Retrieved from https://doi.org/10.1145/3477142 doi: 10.1145/3477142

Hingston, P. (2009). A turing test for computer game bots. *IEEE Transactions on Computational Intelligence and AI in Games*, *1*(3), 169–186.

Hossen, M. I., Tu, Y., Rabby, M. F., Islam, M. N., Cao, H., & Hei, X. (2020, October). An object detection based solver for google's image recaptcha v2. In *23rd international symposium on research in attacks, intrusions and defenses (RAID 2020)* (pp. 269–284). San Sebastian: USENIX Association. Retrieved from https://www.usenix.org/conference/raid2020/presentation/hossen

Iliou, C., Kostoulas, T., Tsikrika, T., Katos, V., Vrochidis, S., & Kompatsiaris, Y. (2019). Towards a framework for detecting advanced web bots. In

Proceedings of the 14th international conference on availability, reliability and security. New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3339252.3339267 doi: 10.1145/3339252.3339267

Jiao, Y., Gabryś, A., Tinchev, G., Putrycz, B., Korzekwa, D., & Klimkov, V. (2021). Universal neural vocoding with parallel wavenet. In *Icassp 2021 - 2021 ieee international conference on acoustics, speech and signal processing (icassp)* (p. 6044-6048). doi: 10.1109/ICASSP39728.2021.9414444

Khatri, C., Hedayatnia, B., Venkatesh, A., Nunn, J., Pan, Y., Liu, Q., ... Prasad, R. (2018). *Advancing the state of the art in open domain dialog systems through the alexa prize.*

Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. *Information Sciences*, *467*, 312-322. Retrieved from https://www.sciencedirect.com/science/article/pii/S0020025518306248 doi: https://doi.org/10.1016/j.ins.2018.08.019

Lebeuf, C., Zagalsky, A., Foucault, M., & Storey, M.-A. (2019). Defining and classifying software bots: A faceted taxonomy. In *Proceedings of the 1st international workshop on bots in software engineering* (p. 1–6). IEEE Press. Retrieved from https://doi.org/10.1109/BotSE.2019.00008 doi: 10.1109/BotSE.2019.00008

Malik, M., Malik, M. K., Mehmood, K., & Makhdoom, I. (2021). Automatic speech recognition: a survey. *Multimedia Tools and Applications*, *80*(6), 9411–9457.

Oord, A., Li, Y., Babuschkin, I., Simonyan, K., Vinyals, O., Kavukcuoglu, K., ... others (2018). Parallel wavenet: Fast high-fidelity speech synthesis. In *International conference on machine learning* (pp. 3918–3926).

Pham, P., Nguyen, L. T., Vo, B., & Yun, U. (2021). Bot2vec: A general approach of intra-community oriented representation learning for bot detection in different types of social networks. *Information Systems*, 101771. Retrieved from https://www.sciencedirect.com/science/article/pii/S0306437921000302 doi: https://doi.org/10.1016/j.is.2021.101771

Prahallad, K. (2010). *Automatic building of synthetic voices from audio books* (PhD dissertation). Carnegie Mellon University.

Radware. (2020). *The big bad bot problem - an analysis of trends in malicious bot attacks and their impact on organizations* (Tech. Rep.). Retrieved from https://www.radwarebotmanager.com/big-bad-bot-problem-report-2020/

Ranoliya, B. R., Raghuwanshi, N., & Singh, S. (2017). Chatbot for university related faqs. In *2017 international conference on advances in computing, communications and informatics (icacci)* (p. 1525-1530). doi: 10.1109/ICACCI.2017.8126057

Rauchfleisch, A., & Kaiser, J. (2020, 10). The false positive problem of automatic bot detection in social science research. *PLOS ONE*, *15*(10), 1-20. Retrieved from https://doi.org/10.1371/journal.pone.0241045 doi: 10.1371/journal.pone.0241045

Sayyadiharikandeh, M., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F.

(2020). Detection of novel social bots by ensembles of specialized classifiers. In *Proceedings of the 29th acm international conference on information and knowledge management* (p. 2725–2732). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3340531.3412698 doi: 10.1145/3340531.3412698

Serban, I. V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., ... Bengio, Y. (2017). *A deep reinforcement learning chatbot.*

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018, November). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), 4787. Retrieved from https://doi.org/10.1038/s41467-018-06930-7

Sivakorn, S., Polakis, I., & Keromytis, A. D. (2016). I am robot:(deep) learning to break semantic image captchas. In *2016 ieee european symposium on security and privacy (euros&p)* (pp. 388–403).

Somanath, G. (2021). Can't find that perfect gift? blame the bots. *Computer Fraud & Security*, *2021*(3), 6-8. Retrieved from https://www.sciencedirect.com/science/article/pii/S1361372321000282 doi: https://doi.org/10.1016/S1361-3723(21)00028-2

TANAKA, T., NIIBORI, H., LI, S., NOMURA, S., KAWASHIMA, H., & TSUDA, K. (2020). Bot detection model using user agent and user behavior for web log analysis. *Procedia Computer Science*, *176*, 1621-1625. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877050920320871 (Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 24th International Conference KES2020) doi: https://doi.org/10.1016/j.procs.2020.09.185

Tardelli, S., Avvenuti, M., Tesconi, M., & Cresci, S. (2021). Detecting inorganic financial campaigns on twitter. *Information Systems*, 101769. Retrieved from https://www.sciencedirect.com/science/article/pii/S0306437921000296 doi: https://doi.org/10.1016/j.is.2021.101769

Wallace, R. S. (2009). The anatomy of a.l.i.c.e. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the turing test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 181–210). Dordrecht: Springer Netherlands. Retrieved from https://doi.org/10.1007/978-1-4020-6710-5_13 doi: 10.1007/978-1-4020-6710-5_13

Wang, Y., Wei, Y., Zhang, M., Liu, Y., & Wang, B. (2021). Make complex captchas simple: A fast text captcha solver based on a small number of samples. *Information Sciences*, *578*, 181-194. Retrieved from https://www.sciencedirect.com/science/article/pii/S0020025521007301 doi: https://doi.org/10.1016/j.ins.2021.07.040

Wirth, K., Menchen-Trevino, E., & Moore, R. T. (2019). Bots by topic: Exploring differences in bot activity by conversation topic. In *Proceedings of the 10th international conference on social media and society* (p. 77–82). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3328529.3328547 doi: 10.1145/3328529.3328547

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for

customer service on social media. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 3506–3510). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025453.3025496 doi: 10.1145/3025453.3025496

Yu, D., Cohn, M., Yang, Y. M., Chen, C.-Y., Wen, W., Zhang, J., ... Yu, Z. (2019). *Gunrock: A social bot for complex and engaging long conversations.*

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, *57*(2), 102025. Retrieved from https://www.sciencedirect.com/science/article/pii/S0306457318306794 doi: https://doi.org/10.1016/j.ipm.2019.03.004