Article

SM+S
social media + society

Social Media + Society
July-September 2020: 1–14

DOI: 10.1177/2056305120939264
journals.sagepub.com/home/sms
SAGE

Demystifying Social Bots: On the Intelligence of Automated Social Media Actors

Dennis Assenmacher[1], Lena Clever[1], Lena Frischlich[1,2], Thorsten Quandt[1], Heike Trautmann[1], and Christian Grimme[1]

Abstract

Recently, *social bots*, (semi-) automatized accounts in social media, gained global attention in the context of public opinion manipulation. Dystopian scenarios like the malicious amplification of topics, the spreading of disinformation, and the manipulation of elections through "opinion machines" created headlines around the globe. As a consequence, much research effort has been put into the classification and detection of social bots. Yet, it is still unclear how easy an average online media user can purchase social bots, which platforms they target, where they originate from, and how sophisticated these bots are. This work provides a much needed new perspective on these questions. By providing insights into the markets of social bots in the clearnet and darknet as well as an exhaustive analysis of freely available software tools for automation during the last decade, we shed light on the availability and capabilities of automated profiles in social media platforms. Our results confirm the increasing importance of social bot technology but also uncover an as yet unknown discrepancy of theoretical and practically achieved artificial intelligence in social bots: while literature reports on a high degree of intelligence for chat bots and assumes the same for social bots, the observed degree of intelligence in social bot implementations is limited. In fact, the overwhelming majority of available services and software are of supportive nature and merely provide modules of automation instead of fully fledged "intelligent" social bots.

Keywords

social media analysis, social bots, misinformation, automated intelligence, human computer interaction

Introduction

Social media platforms like Facebook, Instagram, or Twitter have become an established part of the modern media diet (Newman et al., 2017) and communication nowadays is more often than ever mediated through online technologies. However, not all communication partners in online-interactions are honest interaction partners or even humans. Fully or semi-automated user-accounts, so-called *social bots*, increasingly participate in online-interactions. In contrast to other automated agents (such as web-crawlers or service bots), social bots are designed for one- or many-sided communication and the imitation of human online-behavior (Grimme et al., 2017; Woolley, 2016). The spectrum of (assumed or observed) types of social bots ranges from very simple bots that automate single elements of the communication process (e.g., liking or sharing), over partially human-steered accounts with automated elements (so-called hybrid bots, or "cyborgs", see Chu et al., 2010; Grimme et al., 2017), to autonomously acting agents equipped with artificial intelligence (AI) and learning skills such as Microsofts' Zo[1] or Replika.ai.[2]

Social bots and their influence are heavily discussed in the context of political manipulation and disinformation (Bessi & Ferrara, 2016; Ferrara et al., 2016; Kollanyi et al., 2016; Ross et al., 2019), leading governments across the globe to strive for regulation.[3] Yet, as the detection of social bots remains a challenge, the actual number of social bots and details about their realization remain unclear. Quantitatively, different

[1]University of Münster, Germany
[2]LMU Munich, Germany

Corresponding Author:
Dennis Assenmacher, Department of Information Systems, University of Münster, Leonardo-Campus 3, 48149 Münster, Germany.
Email: dennis.assenmacher@wi.uni-muenster.de

numbers exists: Varol et al. (2017) estimated a fraction of 9%–15% of active Twitter accounts to be social bots, while platforms themselves report on an absolute scale on millions of accounts (Roth & Harvey, 2018). Both indications should be used with caution, as the evaluation of the underlying tools applied for detection have been found to be not sufficiently precise in distinguishing social (spam) bots from other (human) pseudo-users or humans (Cresci et al., 2017; Grimme et al., 2018). Hence, quantitative statements on social bots—relative or absolute—remain as speculation to some extent.

What is clear, however, is that social bots need a technical infrastructure, which can be broadly understood as the combination of (a) the profile on a social media platform and (b) the technical preconditions for partial automation of the account's behavior through the accordant platform's *Application Programming Interface* (*API*) or proprietary mechanisms to interact with the website or app front-end (the latter is virtually equivalent to remote controlling an app or browser). In addition, (c) an algorithmic realization of the account's behavior is needed. In theory, behaviors can range from technically simple amplification tasks such as liking or sharing content, over the automated connection with other accounts, up to creating and posting own content and even interacting autonomously with other accounts. Particularly, these latter "intelligent" bots are those who guide dystopian visions of manipulative social bots. Apart from the multiple theoretical and anecdotal taxonomies, little is known about the relative availability of these different types of social bots. The overall goal of this work is to close this gap by providing a comprehensive analysis of the availability, capabilities, operational scenarios, and implementations of social bot software as well as commodities for sale. From this, it should be derived to what extent automated agents in social and online media have evolved and whether social bots should be understood as intelligent, autonomous acting, and possibly, dangerous players in social networks.

Contribution

This article is the first to provide the results of a comprehensive market analysis of the online sales areas for social bots and related commodities in the so-called *clearnet* and the *darknet* and shows which types of social bots are accessible to each and every Internet user. Moreover, we use an exploratory knowledge discovery approach originally proposed by Kollanyi where we crawl data from open-access social bot development projects shared on the most relevant online-repositories such as GitHub[4] (Kollanyi, 2016). Similar to most of the recent research endeavors, Kollanyi's research focuses on the Twitter platform. Extending Kollanyi's work, we here consider—besides Twitter—all relevant social media platforms and further software collaboration platforms for a comprehensive overview on available software. This analysis allows us to narrow down the availability of specialized social bot technologies for those users with (slightly) higher levels of technological knowledge.[5] Overall, the extensive evaluation of these different data sources provides a broad picture of the evolution and state-of-the-art of social bot technologies and their capabilities. We find that in contrast to the dystopian media narratives, autonomous and intelligently acting "opinion machines" are not easily available, neither for online customers in the clearnet or the darknet, nor for technologically savvy users of collaborative software development platforms.

Structure

The remainder of this paper is organized as follows. First, we review the literature to define the core concepts of intelligence and automation underlying our work and prior research on social bots. Rooted in this state-of-the art, we present an initial investigation of the markets for social bots on (English and German) clearnet and darknet venues. Finally, we provide a comprehensive overview about social bot code on free online code sharing platforms. Taken together, this article provides a comprehensive overview about the social bot ecosystem and the developmental state of social bots, providing much needed empirical insights for the ongoing debate. We review this debate in light of our findings in the final section.

Related Work

The investigation of intelligence in social bots demands some words on the notion of intelligence, the general understanding of bots, and a summary of important related work based on a structured literature review to complement the empirical findings of this article.

Notions of Intelligence and Automation

As stated by many authors and nicely summarized by Legg and Hutter (2007) in a list of popular definitions of intelligence across disciplines, intelligence is still an ambiguous concept. While Boring (1923) defines intelligence in a very indirect (and somewhat circular) way as what is measured by an intelligence test, Bingham (1937) defines intelligence as "the ability of an organism to solve new problems." In a later definition which summarizes much of prior development in the field, Gottfredson (1997) broadened the perspective, stating that

> Intelligence is a very general mental capability that [. . .] involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. [. . .] Rather, it reflects a broader and deeper capability for comprehending our surroundings—catching on, making sense of things, or figuring out what to do. (p. 230)

Although many other perspectives from different disciplines exist, see for example, the *Handbook on Intelligence* edited by Sternberg (2000), some of the aspects stated by Gottfredson continue to shape the aims of AI till today and

are applied in areas such as natural language processing, consulting, theorem proving, robotics, perception, and many others (Nilsson, 2014). We thus adopt Gottfredson's (1997) definition as basis for our work and focus on intelligence as related to reasoning, planning, and learning from context.

Considering intelligence in the context of human–computer interaction, Alan Turing (1950) extensively discussed a communication-level test in his work "Computing Machinery and Intelligence." Using the term "imitation game," Turing defined an experiment to challenge machines for their intelligence; covering the identity of communication partners, can an interrogator identify only by questions and answers whether he or she is talking to a human or a machine? Most interestingly, Turing had already connected the discussion of whether machines could pass this test, with the term of learning machines—an idea, we follow up today in the context of machine learning and AI.

As we deal with bots in this work, we specifically search for AI in automation. The term "bot" finds its origin in "robot" and should be synonymous with automation in this article. Here, a bot is not necessarily a physical robot but may only be represented by a software system.

As motivated earlier, we are especially interested in bots appearing on social media platforms—commonly known as social bots. Social bots gain more and more attention in public discussions and multiple research fields (Ferrara et al., 2016). Broadly, a social bot can be understood as,

> a super-ordinate concept, which summarizes different types of (semi-) automatic agents. These agents are designed to fulfill a specific purpose by means of one- or many-sided communication in online media. (Grimme et al., 2017)

From this perspective, a *chat bot* is a specific type of social bot, focusing on one-to-one communication. Shawar and Atwell (2007) define a chat bot as "a software system, which can interact or 'chat' with a human user in natural language." Commonly, chat bots are designed to help or support human users in specific service situations. Typical use cases are for example customer help desks, telephone answering systems, or service aid for digital education (Grimme et al., 2017). These applications require developing "intelligent" and usable software, able to interact with costumers or pupils.

In other contexts, the purposes of social bot employment might be less innocuous. Particularly in political context, sometimes the term *political bot* is used instead of the term of social bot (Hegelich & Janetzko, 2016). Woolley defines political bots as a subclass of social bots, aiming at the participation in political discussions or the propagation of political opinions (Woolley, 2016). To match humans' expectations regarding legitimate social media users, both political and social bots need to embody human-like behavior, and thus, require a certain degree of intelligence.

Finally, an area of current intelligent bot development is the creation and utilization of *game bots*. Game bots are often applied to help a player being successful in games or operate as automated opponents (Mitterhofer et al., 2009). Typical examples are *farming bots* in online games. These bots perform easy but tedious tasks, like collecting values, while their users are away from their computers (Kadlec et al., 2009). A more general definition is given by Jacobs et al. (2005), introducing game bots as game characters controlled by computers. In contrast to social bots, advanced game bots do not interact with humans on a conversational level but try to provide a challenging adversary (opponent) in the game. In game bot research, a clear trend to increase believable bot interaction is observable (Patel & Hexmoor, 2009). Researchers increasingly apply AI to not only make smarter and more human-like moves (Patel et al., 2011) but also resort to imitating (more) convincing human behavior (Karpov et al., 2013). In game bot development, the integration of AI is a clear and dedicated research stream (Vinyals et al., 2019) and accordingly, the *Turing test* (Turing, 1950) has been modified to fit the behavioral level of game bot interaction (Cornelissen & Grootjen, 2008; Hingston, 2009, 2010). This test does not work on the communication level but on the game play level, which implies that intelligence is also defined on the game play, that is, in another area than intelligence in communicative interaction.

Review on Bot Intelligence

To set a contextual basis for the empirical study, we conducted a structured literature review according to vom Brocke et al. (2009). The goal of the literature review is to (a) elaborate a common sense of the understanding of the intelligence of social bots and (b) to find examples from literature, which discuss the development or evolution of bot intelligence. An investigation of implications and impacts of (intelligent) bots on society is of course strongly connected to the aspect of technological potential (Millimaggi & Daniel, 2019), however, beyond the scope of this work. For an interesting and very interdisciplinary discussion regarding this topic, we refer the reader to a collection of academic essays edited by Gehl and Bakardjieva (2016/2017).

Using Scopus[6] as indicator of the academic debate, we searched for all papers entailing the terms "intelligent" or "intelligence" in conjunction with "chat bot" or "political bot" or "social bot." Overall, 206 papers were identified and augmented by context literature finally covering the time from Alan Turing's "Computing Machinery and Intelligence" up to Cresci (2020).

Since Turing's proposal, the creation of a computer program, which has the ability to imitate human-like behavior, was in the focus of early "chatterbot" development. They

were originally constructed to act within a human–computer conversation setting (text- or speech-based) and to ultimately pass the Turing Test in a predefined scope (Abdul-Kader & Woods, 2015; Jafarpour & Burges, 2010). Earliest implementations, such as the famous ELIZA chat bot, were simple rule-based systems employing predefined sentences from a knowledge database to create sentences based on what the interrogator had said (Weizenbaum, 1966). Often, chat bots are frameworks, composed by different subsystems which all interact with each other to create believable human-like output. The most important component is referred to as the *Responder* (Abdul-Kader & Woods, 2015). Here, the textual output is constructed from given input data. In general, research distinguishes between the following two output creation patterns: *retrieval-based* and *generation-based* (Gao et al., 2019). Retrieval-based systems select suitable output from a given database of possible responses. In contrast, generation-based techniques try to generate the output word-by-word, based on probability distributions. For both creation patterns a variety of techniques were proposed: Statistical approaches based on *Markov Chain Models* (Hutchens & Alder, 1998; Levin et al., 2000), ontology utilization for identifying related concepts, pattern matching and a new markup language, the Artificial Intelligence Mark-up Language (AIML; Abdul-Kader & Woods, 2015; Wallace, 2007). While originally, the main goal of chat bots was to pass the Turing test, the functionality of those programs was utilized in many different domains ranging from social interaction such as flirting or joking (Augello et al., 2008), also for teaching (Kerly et al., 2007; Shaw, 2012) and to supporting people with certain neurological diseases like Parkinson's (Ireland et al., 2015).

With the rise of social platforms in the early 2000s, a new type of bot—the social bot—came into play. As one of the first researchers, Boshmaf et al. (2011) report on social bots as computer steered social network accounts that mimic human behavior, similar to chat bots. They also noted that social bots can appear in groups, called bot nets, and can be employed to spread spam, support politicians, or others by pushing campaigns. According to Boshmaf et al. (2011), the army of simple bots is steered by a so-called *herder*. They describe bots to be able to like or send friend requests and to generate content by retweeting existing content, or crawling the Internet for predefined keywords and tweet the so-identified articles or websites.

Most studies describe the intelligence of social bots as being limited to the purposes of malicious spam attacks or the extraction of user data. Spam attacks are realized by a group of bots, where each bot is able to send requests and post or share predefined content (Zhang et al., 2013). To extract user data, bots typically send out friend requests (Wald et al., 2013). Afterwards, data like birthday dates or phone numbers, can be easily collected by API calls (Boshmaf et al., 2013). This is comparable to the *retrieval-based* approach of the early chat bots, presented before.

During the years, social bots became more and more important as political actors (Hegelich & Janetzko, 2016). A spam bot architecture can be used to spread political opinions or disinformation, for example, during election periods. The malicious use of social bots implied the need for social bot detection strategies. Although the intelligence of social bots seems limited to easy tasks (like sharing predefined content and connect with other users), the detection is complicated (Ferrara et al., 2016), especially, when some "simple" human mimicry, like a day and night rhythm is added. Researchers work on bot detection by applying machine learning approaches (Dewangan & Kaushal, 2016) or reverse engineering strategies (Freitas et al., 2015).

Bot detection is complicated, thus, the current level of intelligence in social bots, that is, their ability to create content oftentimes stays vague. According to the literature, mimicking human behavior is realized by spreading predefined posts, share existing posts, or search for content to share in the Internet (Appling & Briscoe, 2017). Next to external sources, social bots could be enhanced by chat bot technology (Boshmaf et al., 2011). Ferrara reports on a dark market, where customized social bots can be bought to support campaigns on social media (Ferrara, 2017).

With the rise of neural computing and introduction of sophisticated neural architectures such as Recurrent Neural- and Long Short-Term Memory (LSTM) Networks (Hochreiter & Schmidhuber, 1997), researchers started to build more complex chat bot systems which mainly improved the *generation-based* approaches. Microsoft's Xiaoice chat system, which was introduced in 2014, was one of the first architectures that utilized these more sophisticated techniques (Gao et al., 2019; Shum et al., 2018). On 23 March 2016, Microsoft published its "intelligent" chat bot Tay, which utilized Machine and Deep Learning techniques to learn human-like communication patterns based on user-created Twitter data similar to Xiaoice. Unfortunately, the community trained (or hacked) the bot to being racist (Neff & Nagy, 2016). In recent years, further developments were proposed which try to combine retrieval-based and generation-based techniques (Fedorenko et al., 2018). Replika and Microsoft's Zo chat bot are only a few mentions.

While social bot detection mechanisms get more and more sophisticated (e.g., *Deep Q-Learning* and *particle Swarm Optimization* Lingam et al., 2019, or adversarial social bot detection Cresci, 2020), there is not much information about the evolution of social bot intelligence (maybe due to the lack of ground-truth). Yin et al. (2020) propose an approach of intelligent and informative content creation for social bots. The idea is that features from text and image content are extracted and processed through neural networks, so that the bot is able to comment in a rational manner. However, the question if those intelligent types of bots are prevalent on social platforms, stays unanswered.

Only little has been published on the subject of bot technology so far. Notable exceptions are an example (but very

Table 1. Search Terms Used to Determine Clear- and Darknet Markets in Alphabetical Order.

Orchestration	Platform	Account access	Capabilities
Bot	Facebook bot	Fake account	Broadcasting bot
Bot army	Instagram bot	Fake identity	Chat bot
Bot net	Reddit bot	Fake profile	DDos bot
Cyborg bot	Social media bot	Fake user	Fake likes
Political bot	Telegram bot		Fake posts
Propaganda bot	Twitter bot		Fake tweets
Remote bot	YouTube bot		Like bot
Social bot	WhatsApp bot		Phishing bot
Spam bot			Post bot
			Share bot
			Tweet bot

simple) bot code provided by Grimme et al. (2017), the work of Kollanyi (2016), which partially inspired this study, and a recent paper by Millimaggi and Daniel (2019) which analyzes social bot code focusing on Twitter and a small selected set of implementations to extract behavioral patterns that may be harmful.

This study contributes to the literature through both a comprehensive analysis of dark markets as well as an extensive examination of software repositories and strives to conclude on the intelligence potential of present social bots. This may serve as a new building block for discussing the influence of social bots in society.

Clearnet and Darknet Market Places

Data Acquisition

We investigated clearnet and darknet markets in a systematic field study which mirrors the market access of "average" Internet users. Since social bots are highly debated as tools of political manipulation and malicious trading, we focused on the market situation before a large democratic election, namely the German parliamentary election in 2017. Accordingly, the analysis focuses on German and English markets. Markets were identified through a triangulation of the scientific and "grey" literature as well as through special online resources (e.g., darknet related Reddit threats or the now defunct deepdotweb). We included all markets that were online at the time of data collection and were trading at least one item related to social bot technology into the database. A total of $N=97$ clearnet and $N=30$ darknet venues fulfilled our sampling criteria. To account for the much larger share of clearnet venues, we scanned all of the 30 identified darknet markets and 30% of the clearnet markets for relevant commodities using the keywords presented in Table 1. An overview of the examined markets is presented in the Supplementary Material, Table S1.

Data Description

Overall, social bot-related market places and commodities are more prevalent in the clearnet as compared to the darknet. Collecting descriptions of all commodities identified on the few darknet markets offering relevant items ($n=18$ from 30) and a random sample of clearnet markets ($n=30$, corresponding to 30% of all clearnet markets; sampling was performed to reasonably handle manual coding of content) resulted in a database of $n=815$ clearnet and $n=287$ darknet commodities. Extrapolating from the 30% of clearnet markets examined, the share of clearnet commodities was nearly 10 times higher (9.86:1) than the share of darknet traded commodities.

Targeted Social Media Platforms

In both, the clear- and the darknet, social bot-related commodities were available for every larger social media platform and online service. Wide reaching social media sites dominated the market. Of the $n=452$ clearnet commodities that explicitly mentioned a specific platform, 71% referred to Facebook, Instagram, YouTube, or (though much more seldom) Twitter. Similarly, 80% of platform-specific darknet commodities focused on one of these four social media platforms, again with Twitter being the least frequent target of these four. Without going too much into detail at this point (refer to Figure 2), this interesting pattern suggests that paid third-party services inversely correlate with technical accessibility of social media platforms. The more difficult it is to develop social bots for a social platform, the more commodities are available in clear- and darknet.

Capabilities

Slightly more than one-third (36%) of the commodities advertised certain capabilities, allowing insights into the "intelligence" of the offered service or product. Most of them advertised simple amplification functions such as liking or sharing content (20%), followed by—the technically still rather simple—simulation of social connection through following others or providing fake followers (14%). Only 2% of commodities advertised "intelligent" functionality.

A closer inspection of the latter $n=21$ products showed that the "intelligent" functionalities of the three items offered in the darknet were far away from ready-to-use social bots: two were "tutorials for creating chat bots" ($n=2$) and one offered random (i.e., not very credible) Instagram comments. In the clearnet, offering the creation of content (i.e., comments) was more frequent, though still very seldom ($n=18$). Most offered comments for Instagram and YouTube ($n=11$), however, there was also one "auto poster" for Facebook groups, one WhatsApp "spam and traffic injurer," and two tweet bots. As the Twitter bots were tagged as "spammers," their intelligence, however, seems questionable. The last two

commodities referred to instructional material, promising to "teach spamming." Roughly, 10% of all commodities (12% in the darknet, 9% in the clearnet) promised only access to compromised, herded, or faked user-accounts that could later be used within a manipulation scenario by either humans or social bots.

The vast majority (51%) of commodities, however, addressed criminal purposes not directly related to manipulation as discussed in the context of social bots (e.g., spreading Trojans, exploiting SQL vulnerabilities, or conducting DDoS attacks).

Free Online Code Sharing Platforms

Data Acquisition

We used the Alexa global usage statistics to identify the five most relevant collaboration platforms. The Alexa rank is a metric, which can be taken to evaluate the importance of a website.[7] The metric combines calculations of internal homepage traffic such as page callings, and their development over time. Since the collaboration platforms are structured differently, it was infeasible to establish a common and comparable procedure for searching specific bot programs. The largest platform, GitHub, offers a detailed search engine where explicit search criteria can be applied on different repository fields. Similar to GitHub, GitLab also offers an API. We selected the search terms as generic as possible. Concretely, we combined the name of each Social Platform with the term (bot) through a logical AND operator. For GitHub, GitLab, and Bitbucket a unique crawler was programmed that automatically gathered the repositories' information for all search term combinations. While GitHub and GitLab explicitly provide an external API for searching, Bitbucket is not easily accessible. Therefore, we utilized a web scraping framework for collecting the relevant information. The remaining platforms, SourceForge and Launchpad, were manually queried through the provided web interface because of the low number of matching repositories for those platforms.

To allow for time efficient crawling and to avoid noise in the data set due to temporal developments during data collection, we specified the following limitations to our gathering process: First, we did not download the actual files (source code) of the repositories, since our analysis is mainly based on metadata. Second, we dismissed the history of individual commits (code contributions) on all repositories. Although these data may provide interesting insights, the amount of potential additional API requests would have been significantly increased. Instead, we limit our analysis to the first and last contributions. Due to the heterogeneous structure of the collaboration platforms, we defined a common intermediate schema for data representation (see Supplementary Figure S11). Although some platforms consist of meta-data (e.g., location attribute on GitHub) that are not present on other sources, we include these additional information sources in our analysis. This especially holds for the GitHub platform which contains more than 90% of all repositories.

Table 2. Top Five Code Repository Hosting Platforms.

Rank	Code repository	Alexa rank (April 2019)
1	GitHub	45
2	SourceForge	350
3	Bitbucket	770
4	GitLab	1418
5	Launchpad	7778

Data Description

The data of $N=45{,}018$ different code-repositories was gathered from the top five collaboration platforms (Table 2). The database consists of projects created during an entire decade (April 2008 to June 2019). The largest number of repositories was provided by GitHub (43,026), followed by GitLab (1,606), and Bitbucket (386). GitHub clearly dominated the stage. Its competitors, namely GitLab and Bitbucket, hosted only a small fraction of the total number of bot repositories (4%), therefore we considered them as niche platforms.

Targeted Social Media Platforms

Across all collaboration platforms, we observed a similar distribution regarding frequency of social bot repositories targeting specific social media platforms. Most bot code was produced for Telegram (52%), followed by Twitter (24.5%), Facebook (9%), Reddit (9%), and Instagram (2%).[8]

At first sight, this is a surprising result since Telegram is not considered as one of the big social media players and the platform only exists since 2013. A detailed inspection of the creation date for Telegram-oriented repositories revealed that until 2015 the platform did not receive a lot of attention. This changed in June and July 2015, when a significant increase in the number of related projects can be observed. We directly explain this sudden increase by the fact that Telegram officially launched its open bot platform, on 24 June 2015, making it easy for programmers to create automated bot programs through an external API. The reverse effect can be observed for the number of newly created Twitter bot repositories in summer 2018, see Figure 1. Twitter revised the registration process for developer accounts multiple times, by making it successively more complicated to register new applications. While these new restrictions impede the development of new large-scale social bot networks, they also limit the development of sophisticated detection mechanisms which utilize forensic analyses on large amounts of data.

Across all platforms, we observed a positive Spearman rank correlation between the number of repositories found for a specific social platform and the corresponding level of API support ($\rho=.75$).[9] All of the mentioned examples emphasize the crucial

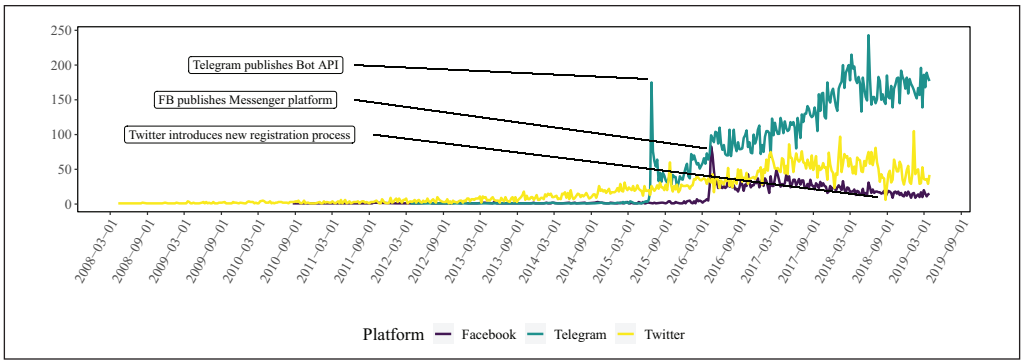


Figure 1. Number of newly created repositories per week.

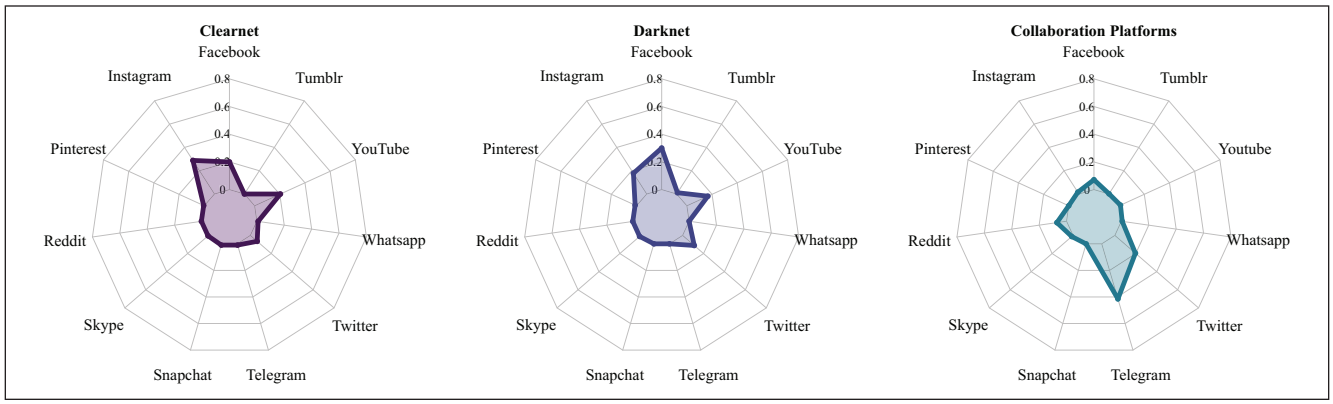


Figure 2. Relative share of bot commodities for the clear/darknet and collaboration platforms.

role of technological affordances for the availability of tools for automation, showing how platform decisions affect the availability and spread of social bots on individual platforms. From an even broader perspective and with respect to addressed social media platforms, we can observe that social bot code availability on open collaboration platforms is contrary to the respective availability on commodities provided in online market places of the clear- and darknet (see Figure 2). This suggests that the level of technical accessibility of social media platforms directly determined whether social bot services were predominantly offered as open software or paid service.

The APIs of the platforms can be accessed by different programming languages. The most common programming language over all platforms is Python, although JavaScript is also frequently utilized. This is most plausible as Facebook explicitly provides a JavaScript Toolkit. In addition, JavaScript is used for accessing the web interface of platforms with restricted APIs (e.g., Instagram).

In 2016, Kollanyi (2016) investigated the geographical origin of Twitter bot repositories. We build on this approach and present an updated version of the contributor distribution in Figure 3. In line with Kollanyi's work, the majority of the Twitter-related repositories originated from the United States. Our updated version reveals that the United Kingdom caught up to Japan and followed the United States by providing the second largest number of bot-related repositories. In addition to the analysis conducted by Kollanyi, we can directly compare the geographical distribution of repositories for all major social media platforms. Therein, we observed some inherent dissimilarities; while Russia did not play an important role in the context of Twitter bots, most of the Telegram bot code owners were from that country, which is presumably due to the popularity of Telegram within the Russian population (Karasz, 2018). For WhatsApp-related bot development, India was the main geographical origin of repositories, as this medium was an important information source and networking platform in that region. Overall, we might infer that the regional social media diet is reflected in the geo-spatial distribution of repositories (Raina et al., 2018).

Bot Capabilities

Identifying the capabilities, operational scenarios, and associated implementation effort of automated bot programs in the context of social media platforms is regarded as one of the major goals of this study.

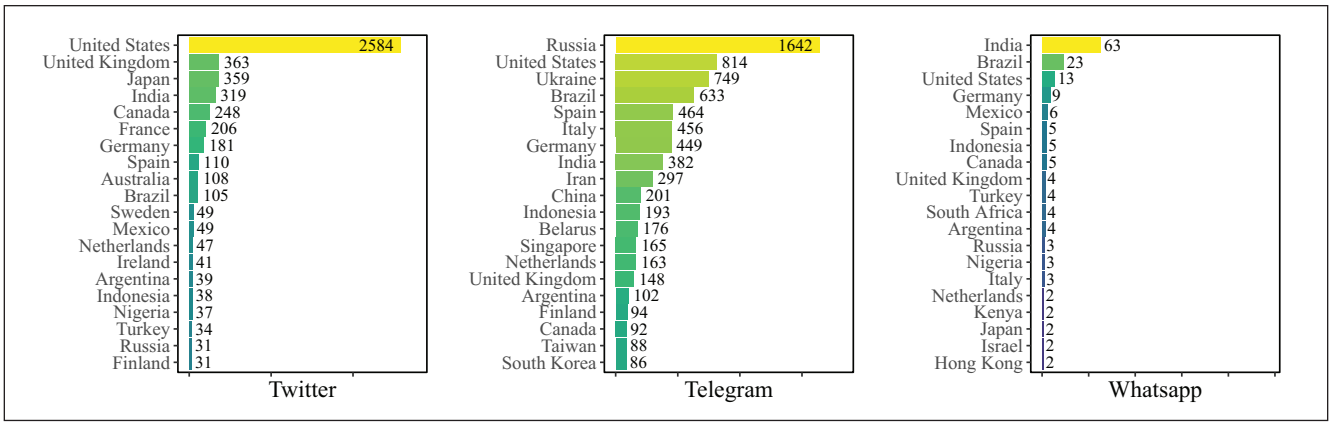
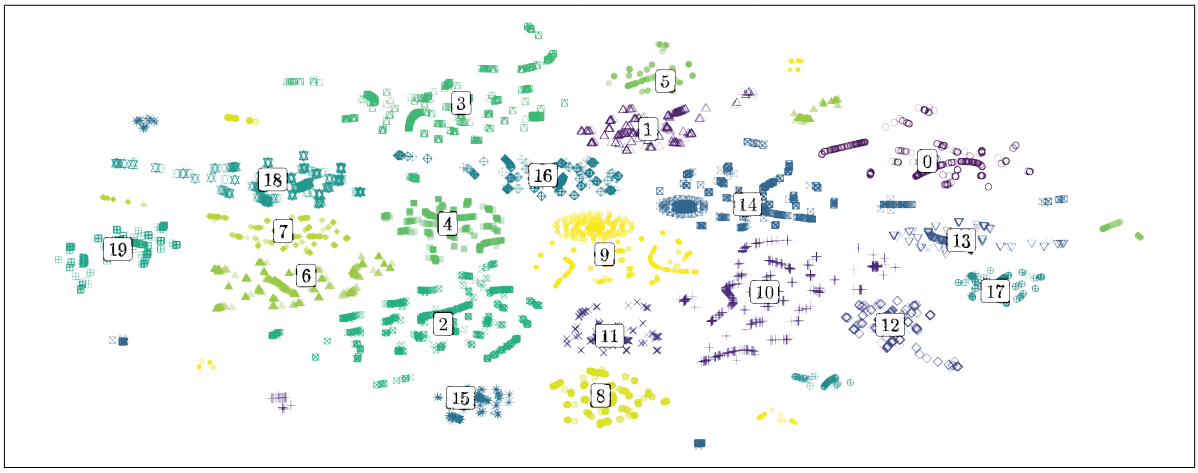


Figure 3. Global geographical distribution of Twitter, Telegram, and WhatsApp repositories determined by evaluating the geographical information provided by the considered software collaboration platforms.

Therefore, the content as well as the overall topics of available social bot code were of central interest. Since, in general, manual inspection of the code base or the description of each repository is infeasible for the number of gathered repositories, we approach the data in two different ways: in a first step, we applied dominance filtering from normative decision theory (Peterson, 2009) to restrict the number of "interesting" repositories for detailed manual investigation. This "interest-value" of repositories was determined through three different indicators: the *number of commits*, the *repository lifespan*, and the *repository timeliness*. While the number of commits is a rough indicator for overall user engagement in a software project, the lifespan (time between creation and last activity) provides insights into the age of the respective repository. The timeliness measures the difference between the current time and the time of the last update. Hence, a low timeliness value refers to a repository that has recently been updated, whereas a high value represents a repository that has not been updated for a long time and is probably inactive. Instead of aggregating these indicators, we identified those repositories that were mutually non-dominating, that is, best compromises, regarding all three indicators. This resulted in 30 repositories to be inspected manually. A large part of the repositories ($n = 11$) contained code, which enabled users to run their own bot script by adjusting a few lines of code. Considering the bot taxonomy of Grimme et al. (2017) those repositories, can be assigned to the class of simple bots. They contained simple API supported functions such as posting, liking, or following. None of the code repositories provided code for "artificially intelligent" acting bots.

Apart from these native social bot functionalities, the largest proportion ($n = 13$) of repositories contained code to handle external API access, realizing code libraries (wrappers) for different programming languages. These repositories can serve as starting points for bot development but they do not provide off-the-shelf executable code for a specific social bot program. The last five code projects were not relevant in context of social bot-creation and provided irrelevant functionality for the scope of this research (e.g., gaming capabilities). Next to our dominance filtering approach, we employed topic modeling to identify homogeneous groups of bot capabilities. In a first step, we selected all repositories with English descriptions of more than one word. In addition, we removed all occurrences of social media platform names from the documents, since the goal of the analysis was the identification of bot code functionality rather than platforms. Next, we applied common data transformation steps such as *tokenization*, stop word removal, and *lemmatization*. In addition to uni-grams, we also create bi-grams to account for close relationships between words within one document, such as word combinations like *Markov chain*. Based on the pre-processed data, we execute Latent Dirichlet allocation (LDA) (Řehůřek & Sojka, 2010). Manual inspection of the results as well as parameter optimization[10] shows that 20 topics were well-suited to reveal existing repository types. Within Figure 4, we exemplary show the word distribution of five of the generated topic distributions together with the most representative repository description. In line with the results of the dominance filtering, the computed topics dealt with simple amplification modules such as spreading image or pre-selected text data (Topics 0, 3). Other topics described projects that utilize frameworks, libraries, or wrappers for different APIs and programming languages to realize automation (Topic 5), while some word distributions represented more sophisticated functionality such as direct user interactions through commenting or chatting (Topics 6, 19). A closer inspection of the related repositories revealed, that even those were of rather simple nature and had limited capabilities. As an example, topic 19 is associated with the implementation of chat components via Markov chains (i.e., creating text-sequences by predicting the next word/token based on the current one). Although basic text-sequences can be generated by those approaches, they often fail to create semantically and orthographically correct statements. Thus, even among these more "intelligent" functionalities, the actual implementations were realized by simple techniques.



t	representative repository description	word distribution
0	Ever wanted to randomly tweet pictures from a folder? No? Well, maybe you do now.	image, random, post, tweet, text
3	A twitter bot that reads in a text file and tweets out all the lines of a text file	tweet, first, create, retweets, go
5	A facebook messenger bot built using flask as a rest API	application, api, created, creating, make
6	Reddit bot that will identify posts with really long walls of text in the self text, then add a comment that splits up the content	comment, post, source, thing, code
19	A Markov chain bot that pulls from Trump's Twitter as well as other sources.	markov, chain, markov chain, article, data

Figure 4. Visualized LDA results and topic distribution of five identified topics. To plot the topics onto a two-dimensional surface, we applied t-SNE (van der Maaten & Hinton, 2008) on the original documents (project descriptions). All documents were represented as 20-dimensional topic distribution vectors and subsequently transformed into a low dimensional space. We colored each document according to the most dominant topic and explicitly excluded descriptions with uncertain topic belonging. The visualization shows that with the chosen number of topics ($N = 20$), we were able to capture homogeneous groups of related repositories. Some wrongly assigned outliers (bottom left) indicate that the model could benefit from a higher number of topics. However, we argue that 20 topics are a good trade-off between interpretability and diversity.

Surprisingly, none of the top representatives among the resulting topics contained concepts related to modern AI or state-of-the-art machine learning algorithms.

Relevance and Methods of Artificial Intelligence

The results of our interest filtering and topic modeling indicate that techniques of AI were rarely utilized and therefore played a minor role in the bot-creation community at the time of data collection. Nonetheless, it is important to identify which techniques are suitable for creating more sophisticated bots. Especially in the aftermath of recent advancements in natural language generation (NLG; Radford et al., 2019; Vaswani et al., 2017), we wanted to evaluate whether these techniques were already part of the established bot-creation tool set or not. To provide a more concrete picture of the

state-of-the-art of intelligent social bots, we manually extracted and analyzed repositories, which contained relevant keywords (e.g., machine learning, deep learning, or AI) or entailed specific algorithms and technologies, for example, *artificial neural nets, long short term memory architectures* (LSTM), or *recurrent neural networks* (RNNs; Hochreiter & Schmidhuber, 1997). A total of 134 distinct repositories (corresponding to 0.3% of all repositories) matched these criteria. A manual inspection of their read-me files as well as the actual program code showed that some of the machine-learning-related repositories did not only contain generative code but also implemented methods for social bot detection ($n = 23$). The projects exclusively employed supervised classification approaches (Hastie et al., 2001) such as support vector machines, logistic regression, or random forests in order to distinguish between human and bot steered accounts. Furthermore, the models were trained on pre-labeled text-corpora, extracted from the Internet (e.g., Kaggle challenges).

A large fraction (25%) of the identified repositories ($n = 32$) was related to chat functionality. Four of these projects did not use machine learning techniques to enable communication (despite their claim). Instead, they only distributed human-generated messages from a message pool on predefined rule sets. Sixteen of the remaining repositories utilized character based RNN/LSTMs which predicted the conditional probability of the next word or token given an already existing sequence. These projects mainly re-implemented online tutorials about the mentioned techniques by training on individual input data. An inspection of the generated samples shows that, similar to Markov chains, the text-sequences failed to generate semantically and orthographically correct texts. Moreover, the projects only focused on sequence generation and did not provide any wrappers to automatically post them on social media platforms. Finally, 12 chat bot projects had a more complex structure. Specifically, the corresponding inspection of the code revealed, that many of these repositories were based on external chat frameworks or libraries,[11] which are able to automatically answer messages. These frameworks were again mainly based on LSTM and RNN structures with more complex architectures and advanced word embeddings. None of the repositories utilized latest state-of-the art models and techniques in the context of NLG such as transformer or *generative adversarial networks* (Subramanian et al., 2018; Vaswani et al., 2017). To create chat bots, the algorithms were usually trained on predefined text-corpora, used as the ground-truth. Corpora either represented content from specific persons (e.g., tweets by Donald Trump, or individual chat histories) or were extracted from large text databases such as novels (e.g., *Alice's adventures in wonderland*). In both cases, the resulting models were only able to imitate the corresponding author style without the ability of generalization. In our perception, none of those generative approaches was able to create convincing human-like content.

Interestingly, we also discovered that $n = 37$ of the $N = 134$ repositories with most promising descriptions (in terms of advertised functionality) contained no code at all or only code fragments with generic code snippets (e.g., using access tokens to fetch data from the Facebook API).

The remaining $n = 42$ repositories contained code for very specific purposes that were not or only indirectly related to social bot functionality. Those features covered, among others, the distribution of cat pictures, execution of sentiment analysis, or detection of jokes. Hence, those projects simply mentioned keywords like "Machine Learning" or "AI" within their description, but did not contain machine learning, deep learning, or AI technologies at all. Only one of the repositories was promoted as an intelligent "stand alone" social bot script, which could be implemented to replace a human user.[12] Yet, when analyzed in detail, this bot offered only simple functionality (as defined by Grimme et al. 2017), entailing, for instance, the ability to filter for interesting content to amplify.

Discussion

The observations of the different descriptive analyses together imply that we can distinguish two scenarios with respect to the effort of social bot development. The first and simple scenario is the one where bot operators aim for trivial usage of the available APIs of social media platforms performing amplifying actions like posting, favoring, or sharing. For these purposes, off-the-shelf software is freely available and feasible. In addition, there is a set of proprietary interfaces and frameworks to easily enable such tasks.

More expensive, though probably still feasible, is the realization of advanced social bots that simulate human behavior on the activity level—not in interaction with others. Openly available social bot frameworks (some even well established and continuously maintained) enabled the enrichment of ready-to-use building blocks with more complex code. Experiments of Grimme et al. (2018, 2017) demonstrated that such extensions are feasible and require only medium technical expertise. It has to be emphasized that even those presumably simple mechanisms can already cause substantial harm to society if they are used in large scale (Ross et al., 2019).

A major gap, however, can be identified between the available open software components and the sometimes postulated existence of intelligently acting bots—that is, bots which are able to produce original content (related to a defined topic), provide reasonable answers to comments, or even discuss topics with other users. The absence of software tools for such (intelligent) tasks can have multiple reasons:

1. The development of those techniques is too difficult and too costly to be provided for the public. This could be a plausible explanation, however, we would then expect software being offered in commodity markets, either in the clearnet or more probably on the darknet platforms. Still, we did not find any of such commodities in our market analysis and thus assume that commodities comprise tools of simple to medium complexity.
2. Existing code may be shared in an obfuscated way. This approach would contradict market mechanisms (the goal of earning money with provided service) as well as the paradigm of software sharing. Hiding potent software products is plausible for secret services but not for the developer community or commodity providers.
3. Techniques for realizing "intelligent" social bots are scientifically too advanced to be subject of current development. It is plausible to assume that social bots are usually applied by groups or persons that do not have sufficient resources nor time for advanced research in AI. It is thus only pragmatic to assume that the costs of hiring human agents who impersonate multiple (fake) accounts is far more time- and cost-effective than developing and controlling intelligent automatons.

Recent reports on experimental intelligent social bots by large software companies, support the last assumption. These reports suggest that there is currently no productive "intelligent" software available (Hempel, 2017; Ohlheiser, 2016). Other observations show that "intelligence" on conversational level is (intentionally) restricted to advanced chat bot capabilities (Stuart-Ulin, 2018), as simple learning approaches are too sensitive regarding external manipulation (Ohlheiser, 2016). As such, the effort for realizing "intelligent" bots can be considered infeasible, today.

Conclusion and Future Perspectives

In this work, we investigated the reported capabilities regarding the intelligence of automated (bot) programs and compared them to existing social bot realizations gathered from code-sharing collaboration platforms as well as to available darknet commodities. Interestingly, we find a clear discrepancy between the theoretic, literature-based, and the practically achieved degree of intelligence. While sophisticated chat bot architectures do exist, which utilize deep learning techniques and are also discussed in literature, we identified a predominance of only rather simple social bot implementations during our data collection. Although, both chat bots and social bots work on the level of human-machine communication, a transfer of artificially intelligent approaches from chat bots toward social or political bots could not be observed.

As already mentioned before in the previous section, we can only speculate about reasons for this discrepancy. It may be plausible that the available artificial intelligent techniques used in chat bots are not sufficient for dealing with highly domain specific content such as political debates. A different explanation could be that intelligent systems are simply not needed to achieve the overall goal, namely manipulation of humans, and that simple amplification systems already satisfy the needs of users that utilize such bots. As the study focused on the descriptive analysis of open bot repositories as well as clear- and darknet market commodities, a clear reasoning about the causal effects cannot be made and should moreover be tackled in an interdisciplinary research endeavor for future work.

Clearly, technologies will change or advance together with research (e.g., new methods in AI) or with modifications in the technological ecosystem of social media networks (e.g., changes in APIs or the accessibility of data). Beyond singular analyses of market places and open software for social bot realization, future research should monitor the bot capabilities and developments over time to track upcoming trends in social bot-creation (Assenmacher et al., 2020). This can point towards the direction of the development of more sophisticated and multifaceted detection techniques, which go beyond account-based, simplistic, and deterministic techniques applied today. The success of these endeavors, however, also depends on the availability of data for research and thus the transparency and responsibility of social media platforms. Besides methodological advances in science, it is up to both politics and platform operators to enable research for keeping track with new development and challenges.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors acknowledge support by the German Federal Ministry of Education and Research (FKZ 16KIS0495K), the Ministry of Culture and Science of the German State of North Rhine-Westphalia (FKZ 005-1709-0001, EFRE- 0801431) and the European Research Center for Information Systems (ERCIS).

ORCID iD

Dennis Assenmacher https://orcid.org/0000-0001-9219-1956

Supplemental Material

Supplemental material for this article is available online.

Notes

1. https://www.zo.ai/, accessed: December, 2019.
2. https://replika.ai/, accessed: December, 2019.

3. For example, the implementation of SB-1001 in California in July 2019, see https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001, accessed: December, 2019.
4. https://github.com/.
5. In our understanding, these users are technologically savvy enough to install and run (often prototypic) software on their own computer and may be able to perform changes in source code.
6. https://www.scopus.com.
7. See Alexa Rank: https://www.alexa.com/, accessed: December, 2019.
8. Other niche platforms (individual share $\leqslant$ 1%) were Skype, YouTube, WhatsApp, Tumblr, Snapchat, and Pinterest.
9. For a detailed list of platform API support, see Supplementary Material, Table S2.
10. The results of the parameter optimization process can be inspected in Supplementary Figure S12.
11. Frequently used frameworks are Fast.ai, Chatterbox, seq2seq, ServentAI, Rasa, textgenrnn, and Watson.
12. https://github.com/nschaetti/pyInstaBot, accessed: December, 2019.

References

Abdul-Kader, S. A., & Woods, D. J. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications*, *6*(7), 72–80.

Appling, S., & Briscoe, E. (2017). The perception of social bots by human and machine. In V. M. Z. Rus (Ed.), *FLAIRS 2017—Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference* (pp. 20-25). Association for the Advancement of Artificial Intelligence.

Assenmacher, D., Adam, L., Trautmann, H., & Grimme, C. (2020). Towards real-time and unsupervised campaign detection in social media. In R. Barták & E. Bell (Eds.), *FLAIRS 2020—Proceedings of the 33rd International Florida Artificial Intelligence Research Society Conference* (pp. 303-306). Association for the Advancement of Artificial Intelligence.

Augello, A., Saccone, G., Gaglio, S., & Pilato, G. (2008, March 4-7). Humorist bot: Bringing computational humour in a chat-bot system. In *Proceedings—CISIS 2008: 2nd International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 703-708). Institute of Electrical and Electronics Engineers.

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US presidential election online discussion. *First Monday*, *21*(11). https://doi.org/10.5210/fm.v21i11.7090

Bingham, W. V. (1937). *Aptitudes and aptitude testing*. Harpers & Brothers.

Boring, G. E. (1923). Intelligence as the tests test it. *New Republic*, *36*, 35-37

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011, December 5-9). The socialbot network: When bots socialize for fame and money. In *Proceedings of the 27th Annual Computer Security Applications Conference*, ACSAC '11 (pp. 93-102). Association for Computing Machinery.

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2013). Design and analysis of a social botnet. *Computer Networks*, *57*(2), 556–578.

Brocke, J. v., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. In ECIS 2009 Proceedings (pp. 2206-2217). Universita di Verona.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010, December 6-10). Who is tweeting on Twitter: Human, bot, or cyborg? In ACSAC'10 *Proceedings of the 26th Annual Computer Security Applications Conference* (pp. 21–30). Association for Computing Machinery.

Cornelissen, A., & Grootjen, F. (2008). A modern Turing test: Bot detection in mmorpgs. In *Belgian/Netherlands Artificial Intelligence Conference* (pp. 49–55). Institute of Electrical and Electronics Engineers.

Cresci, S. (2020). Detecting malicious social bots: Story of a never-ending clash. In C. Grimme, M. Preuss, F. W. Takes, & A. Waldherr (Eds.), *Disinformation in open online media* (pp. 77-88). Springer.

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In *Proceedings of the 26th International Conference on World Wide Web Companion, WWW '17 Companion* (pp. 963–972). International World Wide Web Conferences Steering Committee.

Dewangan, M., & Kaushal, R. (2016). Socialbot: Behavioral analysis and detection. *Communications in Computer and Information Science*, *625*, 450–460.

Fedorenko, D., Smetanin, N., & Rodichev, A. (2018). Avoiding echo-responses in a retrieval-based conversation system. In D. Ustalov, A. Filchenkov, L. Pivovarova, & J. Žižka (Eds.), *Artificial intelligence and natural language* (pp. 91-97). Springer.

Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 french presidential election. *First Monday*, *22*(8). https://doi.org/10.5210/fm.v22i8.8005

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104. https://doi.org/10.1145/2818717

Freitas, C., Benevenuto, F., Ghosh, S., & Veloso, A. (2015). Reverse engineering socialbot infiltration strategies in Twitter. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 25–32). Institute of Electrical and Electronics Engineers.

Gao, J., Galley, M., & Li, L. (2019). Neural approaches to conversational AI. *Foundations and Trends in Information Retrieval*, *13*(2-3), 127–298.

Gehl, R. W., & Bakardjieva, M. (2017). *Socialbots and their friends: Digital media and the automation of sociality* (1st ed.). Routledge. (Original work published 2016)

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*(1), 13–23.

Grimme, C., Assenmacher, D., & Adam, L. (2018). Changing perspectives: Is it sufficient to detect social bots? In G. Meiselwitz (Ed.), *Social computing and social media: User experience and behavior* (pp. 445-461). Springer.

Grimme, C., Preuss, M., Adam, L., & Trautmann, H. (2017). Social bots: Human-like by means of human control? *Big Data*, *5*(4), 279–293.

Hastie, T., Friedman, J. H., & Tibshirani, R. (2001). *Springer series in statistics. The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Hegelich, S., & Janetzko, D. (2016). Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet. In International AAAI Conference on Web and Social Media (pp. 579–582). Association for the Advancement of Artificial Intelligence.

Hempel, J. (2017). *Inside Microsoft's AI comeback*. https://www.wired.com/story/inside-microsofts-ai-comeback/

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

Hingston, P. (2010). A new design for a Turing test for bots. In *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games, CIG2010* (pp. 345–350). Institute of Electrical and Electronics Engineers.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computing*, *9*(8), 1735–1780.

Hutchens, J. L., & Alder, M. D. (1998). Introducing MegaHAL. In *New Methods in Language Processing and Computational Natural Language Learning* (pp. 271-274). Association for Computing Machinery.

Ireland, D., Liddle, J., McBride, S., Ding, H., & Knuepffer, C. (2015). Chat-bots for people with Parkinson's disease: Science fiction or reality? *Studies in Health Technology and Informatics*, *214*, 128–133.

Jacobs, S., Ferrein, A., & Lakemeyer, G. (2005). Controlling unreal tournament 2004 bots with the logic-based action language GOLOG. In AIIDE (pp. 151–152). Association for the Advancement of Artificial Intelligence.

Jafarpour, S., & Burges, C. J. (2010). *Filter, rank, and transfer the knowledge: Learning to chat* (Technical Report MSR-TR-2010-93). Microsoft.

Kadlec, R., Burkert, O., & Brom, C. (2009). Creating game bots in a few easy steps. In *MindTrek 2009—13th International Academic MindTrek Conference: Everyday Life in the Ubiquitous Era* (pp. 222–223). Association for Computing Machinery.

Karasz, P. (2018, May 2). What is Telegram, and why are Iran and Russia trying to ban it? *The New York Times*. https://www.nytimes.com/2018/05/02/world/europe/telegram-iran-russia.html

Karpov, I. V., Schrum, J., & Miikkulainen, R. (2013). Believable bot navigation via playback of human traces. In P. Hingston (Ed.), *Believable bots* (pp. 151-170). Springer.

Kerly, A., Hall, P., & Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems*, *20*(2), 177–185.

Kollanyi, B. (2016). Where do bots come from? An analysis of bot codes shared on GitHub. *International Journal of Communication*, *10*, 4932–4951.

Kollanyi, B., Howard, P. N., & Woolley, S. C. (2016). *Bots and automation over Twitter during the U.S. election* (Technical Report Data Memo 2016.4). Oxford, UK: Project on Computational Propaganda.

Legg, S., & Hutter, M. (2007). A collection of definitions of intelligence. In *Proceedings of the 2007 Conference on Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms: Proceedings of the AGI Workshop 2006* (pp. 17–24). NLD: IOS Press.

Levin, E., Pieraccini, R., & Eckert, W. (2000). A stochastic model of human-machine interaction for learning dialog strategies. *IEEE Transactions on Speech and Audio Processing*, *8*(1), 11–23.

Lingam, G., Rout, R., & Somayajulu, D. (2019). Deep Q-learning and particle swarm optimization for bot detection in online social networks. In *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*. Institute of Electrical and Electronics Engineers.

Millimaggi, A., & Daniel, F. (2019). On Twitter bots behaving badly: Empirical study of code patterns on GitHub. In M. Bakaev, F. Frasincar, & I.-Y. Ko (Eds.), *Web engineering, lecture notes in computer science* (pp. 187-202). Springer.

Mitterhofer, S., Kruegel, C., Kirda, E., & Platzer, C. (2009). Server-side bot detection in massively multiplayer online games. *IEEE Security & Privacy*, *7*(3), 29–36.

Neff, G., & Nagy, P. (2016). Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication*, *10*, 4915–4931.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2017). *Reuters Institute Digital News Report 2017*. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

Nilsson, N. J. (2014). *Principles of artificial intelligence*. Morgan Kaufmann.

Ohlheiser, A. (2016, March 26). Trolls turned Tay, Microsoft's fun millennial AI bot, into a genocidal maniac. *The Washington Post*. https://www.washingtonpost.com/news/the-intersect/wp/2016/03/24/the-internet-turned-tay-microsofts-fun-millennial-ai-bot-into-a-genocidal-maniac

Patel, P., Carver, N., & Rahimi, S. (2011). Tuning computer gaming agents using Q-learning. In *2011 Federated Conference on Computer Science and Information Systems, FedCSIS 2011* (pp. 581–588). Institute of Electrical and Electronics Engineers.

Patel, P., & Hexmoor, H. (2009). Designing BOTs with BDI agents. In *2009 International Symposium on Collaborative Technologies and Systems, CTS* (pp. 180–186). Institute of Electrical and Electronics Engineers.

Peterson, M. (2009). Decisions under ignorance. In M. Peterson (Ed.), *Cambridge introductions to philosophy. An introduction to decision theory* (pp. 40–63). Cambridge University Press.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *Openai Blog*, *1*(8). https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Raina, R. L., Alam, I., & Siddiqui, F. (2018). Facebook losing to WhatsApp: The changing social networking patterns in India. In S. Chhabra (Ed.), *Handbook of research on civic engagement and social change in contemporary society* (pp. 328–346). IGI Global.

Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). European Language Resources Association.

Ross, B., Pilz, L., Cabrera, B., Brachten, F., Neubaum, G., & Stieglitz, S. (2019). Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks. *European Journal of Information Systems*, *28*, 394–412.

Roth, Y., & Harvey, D. (2018). *How Twitter is fighting spam and malicious automation* [Technical report]. Stanford, CA: Twitter.

Shaw, A. (2012). Using chatbots to teach socially intelligent computing principles in introductory computer science courses. In *Proceedings of the 9th International Conference on Information Technology, ITNG 2012* (pp. 850–851). Institute of Electrical and Electronics Engineers.

Shawar, B. A., & Atwell, E. (2007). Different measurement metrics to evaluate a chatbot system. In *Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies* (pp. 89–96). Association for Computing Machinery.

Shum, H.-Y., He, X.-D., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*(1), 10–26.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence*. Cambridge University Press.

Stuart-Ulin, C. R. (2018). *Microsoft's politically correct chatbot is even worse than its racist one*. https://qz.com/1340990/microsofts-politically-correct-chat-bot-is-even-worse-than-its-racist-one/

Subramanian, S., Mudumba, S. R., Sordoni, A., Trischler, A., Courville, A. C., & Pal, C. (2018). Towards text generation with adversarially learned neural outlines. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (pp. 7551-7563). Curran Associates.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. In: *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* (pp. 280–289). Association for the Advancement of Artificial Intelligence.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17* (pp. 6000–6010). Curran Associates.

Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., . . .Silver, D. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, *575*(7782), 350–354.

Wald, R., Khoshgoftaar, T., Napolitano, A., & Sumner, C. (2013). Which users reply to and interact with Twitter social bots. In *Proceedings-International Conference on Tools with Artificial Intelligence, ICTA*I (pp. 135–144). ICTAI.

Wallace, R. S. (2007). 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). Springer.

Weizenbaum, J. (1966). Eliza—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Woolley, S. (2016). Automating power: Social bot interference in global politics. *First Monday*, *21*(4). https://doi.org/10.5210/fm.v21i4.6161

Yin, Y., Wu, H., & Zhang, X. (2020). *Neural visual social comment on image-text content* [IETE Technical Review]. New Delhi, India: Institution of Electronics and Telecommunication Engineers.

Zhang, J., Zhang, R., Zhang, Y., & Yan, G. (2013). On the impact of social botnets for spam distribution and digital-influence manipulation. In *IEEE Conference on Communications and Network Security, CNS 2013* (pp. 46–54). Institute of Electrical and Electronics Engineers.

Author Biographies

Dennis Assenmacher (MSc University of Münster), is a researcher at the Department of Information Systems, University of Münster, Germany. His research interests include social media analytics as well as supervised and unsupervised data stream learning.

Lena Clever (MSc University of Münster), is a researcher of Information Systems at the University of Münster. Her research interests include social media analytics, classification, stream data analysis.

Lena Frischlich (PhD University of Cologne), is currently an interim professor at the LMU Munich, Department of Communication and Media Studies and the principal investigator of the junior research group demoRESILdigital at the University of Muenster, Institute for Communication Science. Her research interests include the staging and effects of online-propaganda and related phenomena.

Thorsten Quandt (PhD Ilmenau University of Technology, habil. Ludwig-Maxmilians-Universität München), is a professor of Communication Studies at the University of Münster, Germany. His research interests include online communication, propaganda and populism, as well as digital games and VR.

Heike Trautmann (PhD, habil. TU Dortmund), is a professor of Statistics and Optimization at the Department of Information Systems, University of Münster, Germany. Her research interests include data science, social media analytics, (evolutionary) optimization, as well as automated algorithm selection and configuration.

Christian Grimme (PhD TU Dortmund University, Habilitation University of Münster), is an associate professor for Information Systems at the University of Münster, Department of Information Systems. His research interests are computational propaganda detection in social media, large scale data analysis, as well as decision support and optimization.