Working Paper No. 2017.4

Computational Propaganda in Poland: False Amplifiers and the Digital Public Sphere

Robert Gorwa, University of Oxford

Table of Contents

Table of Figures

Abstract

This report provides the first overview of political bots, fake accounts, and other false amplifiers in Poland. Based on extensive interviews with political campaign managers, journalists, activists, employees of social media marketing firms, and civil society groups, the report outlines the emergence of Polish digital politics, covering the energetic and hyper-partisan "troll wars", the interaction of hate speech with modern platform algorithms, and the recent effects of "fake news" and various sources of apparent Russian disinformation. The report then explores the production and management of artificial identities on Facebook, Twitter, and other social networks—an industry confirmed to be active in Poland—and assesses how they can be deployed for both political and commercial purposes. The quantitative portion of the report features an analysis of Polish Twitter data, and demonstrates that a very small number of suspected bot accounts are responsible for a disproportionally large proportion of activity on the sampled political hashtags. Furthermore, within this dataset, there appear to be twice as many suspected right-wing bot accounts as there are left-wing accounts. These right-wing accounts are far more prolific than their left-wing counterparts, with a tiny number of highly active right-wing accounts generating more than 20% of the total volume of political Twitter activity collected over a three-week period. Overall, the report provides evidence for a rich array of digital tools that are increasingly being used by various actors to exert influence over Polish politics and public life.

Introduction

Since the 2016 US Election, an increasing amount of public attention has been paid to the effect that digital disinformation is having on democracy and political life in the West. Leading newspapers, captivated by the apparent influx of "fake news" and the various online influence operations that seem to have targeted political campaigns in countries such as France and the United States, have in recent months covered bots, trolls, and various other, previously esoteric aspects of the digital public sphere. In a sense, this was to be expected: as the online dimension of politics became more prominent, so did the likelihood that efforts to shape online media ecosystems and manipulate public opinion on social networks would emerge (Woolley & Howard, 2016). A recent body of scholarship has begun to engage with the various new forms of "computational propaganda," such as automated social media bots, organized networks of fake online identities, and coordinated trolling campaigns that have become increasingly prevalent and are rapidly being established as an important aspect of contemporary digital politics (Woolley, 2016). However, scholarly understanding of these developments remains limited, especially in countries outside of Western Europe and North America. For all the talk of bots, trolls, and "fake news" in the United States and United Kingdom, it is not entirely clear if they pose an issue elsewhere. Have these phenomena spread? And if so, how are they understood and perceived in other countries?

Poland provides a fascinating case study for a variety of reasons. Firstly, despite the numerous cases of alleged political trolling and online manipulation by foreign actors that have been covered in the Polish media, as well as a highly adversarial domestic political climate and accusations that certain Polish political parties are using paid commentators and fake accounts on a variety of social networks, there have been no comprehensive efforts to assess these developments in the country. Secondly, Poles have in recent years eagerly embraced multiple online platforms, and today the Internet has become very important for political life in the country. In particular, Facebook has emerged as a central source of political information and news, and is perhaps even more influential in Poland than it is in countries like the United States, at least for younger users. Finally, Poland's complex history and current political climate combine to yield a challenging yet unique environment for any study.

This report aims to provide an initial exploration of computational propaganda and media manipulation in Poland, and in the process, shed further insight into the general operation and effects of bots, fake accounts, and other false amplifiers.

It proceeds in six parts. In the section that follows, key terms are defined and the report's methodology is discussed. In the third section, background for the case study is provided, and various recent developments in Polish digital politics are discussed, including the energetic and hyper-partisan "troll wars", the interaction of hate speech with modern platform algorithms, and the influence of "fake news". The fourth section discusses the various sources of apparent Russian disinformation to which Poles are regularly exposed to, as well as what is believed to be Russian-linked activity on Polish social networks that has persisted since the onset of the 2013 Ukraine Crisis. The fifth section explores the production and management of artificial identities on Facebook by Polish political consultancies and social media marketing firms, and assesses how they can be deployed for both political and commercial purposes. The final section outlines four improved heuristics for flagging suspected bot accounts and uses them to perform an analysis of Polish Twitter data.

Definitions and Methods

Setting baseline definitions for the processes being observed allows one to better understand how the observations from our study adhere to, or deviate from, the commonly held conceptions of these phenomena. As we will see, these definitions can be flexible and are often contested.

Howard and Woolley have theorized that three main elements—political bots, organized trolling campaigns of hate and harassment, and the online dissemination of 'fake news' and disinformation—form a broader system of *computational propaganda,* an "assemblage of social media platforms, autonomous agents, and big data tasked with the manipulation of public opinion" (Woolley & Howard, 2016, p. 4887). These are explored in turn.

Bots

Shortly following the emergence of Twitter as a major microblogging service in the late 2000's, certain computer scientists began to express interest in *social bots*, automated accounts that mimic users on social media platforms (Lee et al., 2011). Scholars noted that Twitter's fairly open API was conducive to its flexible integration with many apps and third-party services, but also made it quite easy for bots to proliferate, leading some to suggest that this increase in automation could create a "double edged sword" for the platform, as benevolent bots would inflate Twitter's user numbers and "generate large numbers of benign tweets," while also allowing for the possibility that more malicious bots could manipulate hashtags and spread spam (Chu et al., 2010, p. 21).

Most recently, social scientists have become concerned about the influence of partisan *political bots*, especially in the run up to major elections (Howard & Kollanyi, 2016; Woolley, 2016). In the simplest sense, these are bots that serve some political function, and political bots are generally, but not always, *social media bots* (bots that operate on social media), designed to "mimic real people so as to manipulate public opinion across a diverse range of social media and device networks" (Woolley & Howard, 2016, p. 4886).

There are many different types of bots, performing a variety of tasks online. For example, Tsvetkova and colleagues outline the many different types of bots that tend to perform one or more of four broad functions: they can collect information,

execute actions, generate content, and emulate humans (Tsvetkova et al., 2017). These bots can be benign—for example, there have been several examples of Twitter bots that attempt to foster positive online discourse—but more malevolent bots also exist, spreading spam and malicious links (Murthy et al., 2016; Ferrera et al., 2016). Exactly how much automation is required for an account to be properly considered a bot is still an open question, but for the purposes of this paper, *bot* simply refers to an automated account on an online platform.

Trolling and Fake Accounts

Another increasingly important element of political life online is *trolling.* Trolling is difficult to define and has its roots in the early days of bulletin boards such as Usenet (Coleman, 2012). As Marwick and Lewis (2017, p. 4) note, the term initially "described those who deliberately baited people to elicit an emotional response". But since the early 2000s, scholars have demonstrated how playful trolling emerged on certain online forums but eventually would become more synonymous with hate and harassment as demonstrated on message boards such as 4Chan's /b/ (Herring et al., 2002; Marwick & Lewis, 2017). While key questions about trolling today remain unanswered, elements of trolling have been established as an important aspect of twenty-first century online political mobilization (Beyer, 2014).

In the past few years, investigative journalists have shed light on different forms of government sponsored or organized activity on a variety of social networks, with Adrian Chen most notably investigating a Russian operation in St. Petersburg that was allegedly home to hundreds of employees paid to post comments on articles, write blog posts, and attempt to influence political debates on social media in a variety of ways (Chen, 2015). This kind of operation is commonly called a "troll-farm" or "troll-army" by commentators, although it does not ascribe to traditionally held definitions of what constitutes trolling and possibly should not be classified as such. Others have called these sorts of users *sockpuppets* (Woolley, 2016, p. 4), but for the sake of clarity, this paper will refer to fake accounts on Facebook or other platforms simply as "fake accounts".

'Fake News'

Finally, "fake news" has become an especially popular term in recent months (Allcott & Gentzkow, 2017). However, as it has come to mean everything from tabloid "clickbait" content to overt misinformation, and seems to have been recently subverted by Donald Trump and the American alt-right media, it is a

particularly difficult concept for researchers to operationalize (Starbird, 2017). For the purposes of this paper, "fake news" will generally be referred to as meaning intentionally incorrect or misleading information spread by a news organization (real or not) for political purposes.

Methodology

This study was conducted using a mix of qualitative and quantitative methods. The qualitative portion consisted of ten semi-structured and anonymous interviews conducted in Poland. Interviews were selected with a hybrid purposive/snowball sampling strategy, where potentially interesting political campaign managers, journalists, activists, employees of social media marketing firms, and digitally minded civil society members were sought out and asked to recommend further interviewees. Interviewing has been shown to be one of the best currently known methods for understanding computational propaganda, given the difficulties inherent in studying processes which often occur behind the scenes on social media platforms that do not share data with researchers (Woolley & Howard, 2016). These interviews were further informed by approximately two dozen informal and off-the record conversations with a variety of Polish experts. In conjunction to these interviews, a study of Polish Twitter was undertaken together with Bence Kollyani and the Computational Propaganda Research team. The methodology for that element of this report will be discussed in the sixth section, "Automated Accounts on Twitter".

Background: The Emergence of Polish Online Politics

In 1991, the first Polish Internet connection was established between the University of Copenhagen and the University of Warsaw (Trammell et al., 2006). After dial-up Internet access became widely available in the country in 1996, various forms of online communication, such as bulletin boards, emerged and would grow steadily, eventually being supplanted by early blogging platforms (Trammell et al., 2006). These set the stage for the first Polish social network, *NaszaKlasa* ("Our Class"), which was launched in 2006 by a group of university students from Warsaw. Designed as a method for classmates stay in touch after graduation, it became a popular platform and experienced impressive growth in the late 2000s. In the past few years, however, Poles have increasingly shifted towards a variety of next generation online platforms and forums (Koc-Michalska et al., 2014). In 2011, the overall Internet penetration rate was around 59%, and there were only 5.5 million

Polish Facebook users, but in the past six years, household Internet penetration is said to have increased substantially to 80%, for of a total of approximately 30.4 million total Internet users (Eurobarometer, 2016). According to the most recent data available, more than three quarters of those online are now on Facebook, which now has approximately 22.6 million users in the country (Gemius/PBI, 2017).

As these numbers continue to grow, Polish academics have begun to engage with the ways the Internet and social media platforms are affecting political communication in the country. Specifically, scholars have noted the steadily increasing importance of the Internet as a vehicle for political marketing in Poland (Baranowski, 2015). Since the 2011 Federal election—held up as the first time that the Internet was used broadly by candidates from multiple parties—campaigns have been using an increasingly professionalized set of tools to manage their online self-presentation and mobilize supporters (Koc-Michalska et al., 2014). These include various social networks and the online marketing tools that can be deployed on them. Now, many politicians have a visible Twitter presence, although Twitter is still widely seen as an 'elite' platform for journalists and politicians (Baranowski, 2015). As of 2015, there were 4 million Polish Twitter users (Sotrender, 2016b).

Factoring into these shifts is Poland's unique political situation. Only a few years ago, Poland was being praised as the premier example of a thriving post-soviet democracy (Simons, 2008). In the past several years, however, the political climate has changed substantially, with the governing Law and Justice (*Prawo i Sprawiedliwość*, abbreviated as PiS) party having set off a series of constitutional crises after its rise to power in the 2015 federal elections. Poland's new government has drawn international condemnation for measures said to limit freedom of expression, and triggered a series of highly publicized mass protests on multiple political and social issues (Kublik, 2016; Rankin & Traynor, 2016). While commentators have tended to regretfully chalk up these shifts to the broader recent trend of right-wing populism in Europe, an interesting aspect of these changes that has largely remained under explored is the role that may have been played by the Internet and social media.

Since the 2015 election, journalists and commentators have reflected upon whether PiS "won the Internet" during its successful campaign (Głowacki, 2015). The broad consensus seems to be that PiS managed to mobilize their supporters and control media narratives far more effectively than its opponents, the Civic Platform party

(*Platforma Obywatelska*, abbreviated as PO). This is surprising because PiS's traditional demographic base is generally older and more rural than its competitors' (and is not traditionally conceived as a particularly Internet savvy audience). Some have gone as far as to suggest that PiS's ability to successfully engage and convert young people was a key, if not the key, factor for its success (Dubinski, 2015). As younger Poles rely on digital news sources and social networks for their political information, the various forces shaping online politics in the country have become increasingly important. Some of these phenomena (such as trolling, "fake news", Russian disinformation, fake accounts, and social media bots) are briefly explored in the following three sections.

Trolling, Activists, and Civil Society

Facebook is the most important social network and by extension, the most popular online space for online political debate and discussion. It has in recent years become a highly energetic political forum, and at least as early as 2014, networks of Antifa (meaning anti-fascist) groups have clashed with far-right groups on Facebook, using mass flagging and reporting to pull down their Facebook pages and ban users. According to one interviewee, a political activist, the golden era of these flagging wars (or "troll wars") was in late 2014 and early 2015, when left-wing groups were successful in blocking the pages of many right-wing groups (Tinker, personal correspondence, 22/12/17). In late 2016, this issue once again came to the fore when the Facebook pages of several prominent Polish nationalist groups were blocked, some of which had hundreds of thousands of likes (Woźnicki, 2016). This seems to have been part of a massive flagging campaign organized by several left-wing Facebook groups a few weeks before a controversial nationalist parade in Warsaw.

One such group, with a Facebook page titled the "Organization for Monitoring Racist and Xenophobic Behaviour" proclaimed its victory, claiming responsibility for the bans and saying that these bans were important because they would cut off the Facebook advertising revenue stream for these pages before signing off with "good night white pride" (see *Appendix A*). Facebook reinstated the pages after government pressure, but the incident has sparked conversations about freedom of speech online and demonstrates the ways in which groups of online users have organized online to successfully make high-profile political statements (Urbanek, 2016).

Another major source of political, commercial, and social information for Poles are online-only news 'portals' such as ONET, and Virtual Poland.[1] These are basically online news sites, but feature cross-platform integration and sections for comments and discussion, and according to Alexa, are the two most popular news websites in the country (Alexa, 2017). All of these platforms are now political, and trolling on these websites has become increasingly prevalent. The problem of political trolling and spam on comment has gotten so pervasive that the comment sections on several news sites, most notably the premier Polish weekly, *Gazeta Wyborcza*, have been modified to make it more difficult for users to reply to each other (Sobkowicz & Sobkowicz, 2012). Another commonly reported rumour is that political parties may have been paying users to comment on articles on certain platforms (Wieliński, 2015).

However, activists and journalists in Poland do not have conclusive evidence that this trolling is automated or centrally organized, but several interviewees suggested that Polish right-wing and nationalist groups were mobilizing online in a highly effective way that seems to combine new and traditional modes of organization. By leveraging traditional mobilization networks, such as the youth organizations that have been long associated with various political parties, as well as emailing lists, closed Facebook groups, and group WhatsApp chats, a group can issue specific instructions to its supporters as to what content they should share, where they should comment, and how they can best steer online discussion on key issues. The general lack of neutral online platforms for debate on Polish politics (Sobkowicz & Sobkowicz, 2012) has allowed energetic groups of supporters to infiltrate and spam the comment sections and forums occupied by their clearly defined political opposites. Activists are particularly likely to be caught in the crossfire, especially those that become visible in the public media. "Trolling is an everyday thing", said one digital-rights advocate, "All activists know it is a part of their life now" (Bentham, personal correspondence, 14/02/17).

Even in Poland, emerging forces of trolling and hate speech are interacting with an online experience that is increasingly governed by algorithms, with various interesting and troubling effects. In one notable example, a journalist writing in a prominent publication was "outed" by mocking users posting in the comment section. Although these comments were promptly deleted by moderators, they

[1] *Onet.pl*, *wirtualnapolska.pl*

were online long enough to be picked up by Google's indexing algorithm, and searches of the journalist's name would suggest embarrassing autocomplete results that were supposed to be private (e.g. those searching for "John Doe" would see "John Doe is gay" as the top suggestion).

With the help of a Polish digital rights NGO, the journalist took his case to Google, which initially argued that it could not affect the autocomplete results as they were algorithmically generated, but eventually agreed to change them (Głowacka et al., 2016). This presented itself as a fascinating "Right to be Forgotten" case, as the central issue was not with online content itself, but rather with algorithmically generated tags that were automatically attached to this content. In the words of one interviewee, this example shows that in the age of algorithms, "trolling and hate can generate lasting effects" that may not be immediately apparent (Esme, personal correspondence, 17/02/17).

Fake News

Much like the rest of the world, Poland has recently been seized with the apparent emergence of "fake news." As elsewhere, the phenomenon is still not particularly well understood, although commentators and even major Polish television shows have run exposés touching on this issue. In a few cases, hoaxes and unsubstantiated information spread online in other countries have made it into Poland. For example, the Polish Ministry of Education recently sent out a letter to all schools warning of a social-media based suicide game called "Blue Whale" (*Niebieski Wieloryb*) that had apparently already led to the death of dozens of teenagers in Eastern Europe. However, the story was shortly thereafter revealed to be a hoax, originating on a Russian news site before being reprinted by the English *Sun* newspaper and getting picked up by Polish outlets (Napiórkowski, 2017). There have yet to be explicit examples of political hoaxes and fake news that attain this same level of reach, but the propagation of *fejki* (fakes) and other forms of disinformation has become a prominent concern for many Polish commentators.

It is important to note that Poland has long had a complex media climate, one that may be unique among former Warsaw Pact countries (Pfetsch & Voltmer, 2012). Even during the Communist days, a strong civil society and widespread *samizdat* (underground press) literature spread independent and opposing ideas, factors which led scholars to predict that Poland's diverse media climate would prove

highly resistant to political maneuvering (Pfetsch & Voltmer, 2012). However, this narrative has been challenged in recent years, as political parties have in the past decade done their best to exert their influence over the general media climate. The Law and Justice party (PiS) drew widespread condemnation in both Poland and the West after passing controversial media reform laws that give it more influence over the state-backed broadcaster, TVP, which is now widely seen on the left as an official channel for PiS propaganda. However, it has been pointed out that the previous governments, including the Civic Platform government that was in power earlier, similarly passed policies that intensified the polarization of the Polish traditional media. This underlies the especial difficulties of understanding "fake news" in a country like Poland. One research subject, an academic who studies Polish social media, stated that it is incredibly challenging to meaningfully study "fake news" when the state-backed television channel, TVP, has repeatedly been shown to itself be propagating objectively false information, and when media outlets are viewed as inherently partisan in some way or another (Miller, personal correspondence, 17/02/17).

In sum, the networked public sphere in Poland has grown considerably in the past decade, and a variety of political forces have combined to make Polish online politics energetic, partisan, and often controversial.

Russian Disinformation and Fake Accounts

Along with these domestic forces, Polish online politics have unquestionably been affected by recent events in Ukraine and the complicated Polish-Russian and Polish-Ukrainian relationships. As Polish officials had spent more than two years pushing for deepening ties with Ukraine and were supporting Ukraine's European aspirations, they were naturally troubled when the Ukrainian President, Viktor Yanukovych, chose not to sign the Ukraine–European Union Association Agreement in November of 2013, sparking massive protests and the Ukraine crisis (Przełomiec, 2014). This moment has widely been pointed to as the beginning of what is often perceived to be an active campaign of Russian disinformation propagated via Polish social networks.

As Russia is rumoured to be actively funding nationalist groups, spreading propaganda online, and using other indirect means to destabilize the Polish state, the notion that Russia is engaging in "information operations" or an "information

war" in Poland has become quite popular amongst Polish scholars and commentators, and come to have dominated recent work on propaganda in Poland (Nimmo, 2015; Ostrowki & Woycicki, 2016). A recent report published by the Warsaw-based foreign policy think-tank, the Centre for International Relations *(Centrum Stosunków Międzynarodowych)*, titled "Exposing and Countering pro-Kremlin Disinformation in the Central and Eastern European Countries" provides a series of typical examples. It argues that a variety of dubious outlets spread false information in an effort to undermine the NATO Summit held in Warsaw in the summer of 2016 (Wierzejski, 2016). From fabricated interviews with high-ranking Polish military leaders to sensational attempts to stir up Polish-Ukrainian tensions, the report cites multiple cases in which anonymous "journalists" and bloggers, believed to be linked to Russia, published dubious information that was spread on Facebook and Twitter.

The report notes that this information has occasionally trickled into the mainstream press and has been picked up by large Polish news organizations (an example being when TVP reported a false story about Egypt selling Russian warships that had been originally shared by a questionable Russian news site). Furthermore, the report claims that "Russian trolls are very active in Poland", and relies on manual heuristics (such as poor Polish grammar and the use of Russian idioms) to claim that Russian fake accounts are common on the biggest Polish news portals (Wierzejski, 2016, p. 3). However, as concrete attribution of certain accounts and stories directly to Russian agents is usually impossible, the report is not able truly provide conclusive evidence for its claims. In a bizarre twist that illustrates the complexities of today's online disinformation ecosystem, Sputnik.pl, the Polish branch of Russia's controversial Sputnik News Agency, critiqued and mocked the report's methods in a satirical polish-language article titled "How to Spot a Russian Troll" (Sputnik Polska, 2017).

Despite the protestations of Sputnik, there is considerable circumstantial evidence that indicates that a few days after the Euromaidan protests broke out in Kiev, large numbers of fake accounts flooded Polish Facebook and news portals to weigh in on debates related to Ukraine (Savytsky, 2016; Szczepaniak & Szczygieł, 2017). According to one interviewee, a journalist working on the Caucasus and Eastern European issues, most online discussions touching on Russia held in an open online forum or public Facebook group would quickly be targeted by accounts that spammed comment sections and insulted or harassed commentators.

Those brave enough to engage in discussion on the topic of Russian-Ukranian-Polish relations under their real name would face the threat of targeted hate and harassment. This seems to have become particularly common for journalists and other civil society members, with one interviewee noting that although he had gotten used to the spam and harassment that he would receive after he published articles critical to Russia, it became particularly worrisome when he began receiving private Facebook messages from anonymous accounts that threatened his wife and children by name. Journalists who attempt to engage with these commentators on the portals themselves (or expose them as potentially fake accounts) are especially likely to receive threats and insults (see *Appendix B*).

A 2015 report published by the Polish Government's Computer Emergency Response Team noted Russian influence in Polish cyberspace, and especially on Polish social networks, as a prominent concern (CERT Poland, 2015). However, determining what precisely constitutes Russian influence (or Russian trolling) is a difficult matter: when it comes to conventional cyber activity, attribution is difficult, but governments maintain various investigative options (Rid & Buchanan, 2015). However, the nature of modern disinformation campaigns, especially those conducted via fake accounts, is that they are effectively impossible to conclusively attribute to a certain actor.

While it may have once been possible to identify suspect accounts via certain manual heuristics (for example: the number of friends, choice of profile picture, the use of Russian figures of speech or spelling), evidence suggests that in the past few years it has become significantly more difficult to do so, especially on non-transparent platforms such as Facebook. As one interviewee (a researcher working at a think-tank that deals with cyber issues and attempts to map and track fake Russian accounts) noted, suspected Russian accounts on Facebook have been steadily increasing in their sophistication and seem to now feature more believable profile photos and larger networks of friends. While everyone seems to suspect that Russian-linked organizations or actors are using large numbers of fake accounts on Polish social media platforms, nobody has managed to find evidence or concrete data at a broader level.

Many have attempted to infer the broader goal or motive behind these apparent Russian campaigns. Some have speculated that the goal is to undermine a

population's trust in institutions, spread conspiracy theories, and discredit the idea of truth itself (Pomarantsev & Weiss, 2014). In Poland specifically, others have argued that, "Kremlin narratives seek, paradoxically, to promote extreme Polish nationalism—even anti-Russian nationalism—with the goal of making Poland seem unreliable and 'hysterical' to its Western allies" (Ostrowki & Woycicki, 2016). The combination of fake accounts, fake news sources, and targeted narratives propagated via social media is increasingly becoming portrayed as a new form of digital propaganda. But Polish researchers face two problems: the first is with determining what exactly should be considered propaganda, as it is a politicized term and carries an inherent value judgement.

Should pro-government propaganda be treated the same as propaganda that is apparently foreign in origin? The second is with attributing this propaganda to a specific actor, and trying to meaningfully assess its effects. In the short, medium, and long term, do users really have their opinions changed when repeatedly exposed to these narratives online? Research is sorely needed into this matter. However, a point can be reached where the political discourse becomes saturated to the point that determining true causation may be less important. One research subject memorably noted that "it does not matter if the Russians are actually using fake accounts or bots" to influence online debate in Poland, "as either way, they have succeeded in poisoning the political discourse". They suggested that calling someone a "Russian bot" was rapidly becoming a new slur, deployed to discredit any opinion that was not completely hawkish on Russian affairs. If Poles have begun to constantly accuse each other of being Russian agents if they express unpopular opinions, this is a significant development, and one that does not bode well for the health of online political discourse in the country.

The New Age of Political Marketing: An Insider View

It is interesting to note that the term bot seems to have a different connotation in Poland than in the United States or United Kingdom. As opposed to having a conception of a bot as some kind of script or automated agent, interviewees seemed to broadly view bots as synonymous with trolls. From this perspective, an account would be a bot in the sense that they are seen to be a cog in the Russian propaganda machine (a Russian "bot"), regardless of whether they are operated by a human user or a simple algorithm.

This may be because fully automated social bots, as commonly seen Twitter in the US, were perceived by the interviewees as relatively uncommon on Polish Twitter. The bigger concern seemed to be with what are often termed "troll-farms", networks of fake accounts on social media platforms that are manual (and still predominantly backed by a human user). And it is not just foreign fake accounts (be they real or perceived) that are a source of public concern, as Polish political parties are rumoured to be active in this space as well. Multiple journalists and politicians have accused PiS of using paid "haters" or "trolls" on social media platforms and news portals as part of their extraordinarily effective online resurgence (Głowacki, 2015).

On Twitter, suspicious accounts with no profile photos that engage with other users on political issues have been termed "Szefernaker's Eggs" after Paweł Szefernaker, a Secretary of State in the Chancellery of the Polish Prime Minister who has been referred to as PiS' "internet genius" and is widely believed to be the mastermind behind its successful online efforts (Krzymowski, 2016). While journalists and commentators have investigated some of these operations with varying degrees of success, and there is a great deal of speculation as to how these sorts of operations work, relatively little is known about how these techniques in practice.

Valuable insight into the nebulous underground ecosystem of false amplifiers was provided on the condition of anonymity by a research subject who is a political consultant and marketer, and works for a communications firm that has experience in using fake identities on Polish social media platforms. Over the past ten years, his firm (which we'll refer to here as "The Firm") created more than 40 thousand unique identities, each with multiple accounts on various social media platforms and portals, a unique IP address, and even its own personality, forming a universe of several hundred thousand specific fake accounts that have been used in Polish politics and multiple elections (Daedalus, personal correspondence, 14/01/17).

The process begins with a client: a company in the private sector (pharmaceuticals, natural resources), or a political party/campaign. A strategic objective is outlined and a contract that includes "word of mouth" or "guerrilla" marketing services is written up. An employee of The Firm then starts by creating an email address via a large provider (such as Gmail). Using this email and an invented name, they create accounts on multiple platforms and portals. A suitable profile photo is found via an image search and modified in Photoshop so that it will not appear in a Google

image search, and the employee begins posting on various platforms and building a comment history. Each employee manages up to 15 identities at a time, with each having a coherent writing style, interests, and personality. They use a modified VPN to spoof IP addresses so that their accounts will have a series of associated addresses, allowing them to post from multiple locations in a predictable way (as would befit a normal user using a mobile phone and travelling around a city, or using their laptop from home/work/elsewhere).

When these accounts are ready to begin posting on comment sections and Facebook groups or pages, the employee uses only unique content (each account never copies or repopulates posts) as to make it unsearchable and difficult to link to other accounts. All steps are taken so that these accounts are very difficult (in the words of the research subject, "completely impossible") to conclusively identify as fake.

This all provides a level of deniability for the client, who may not even know exactly (and probably does not want to know) what techniques are being used by their marketing consultants. Furthermore, this is a low risk endeavor: while these processes violate the terms of service for platforms, they exist in a legal grey area. If a firm takes the basic precautions described above, it is highly unlikely that this activity will ever be exposed, and if it is, it is not clear how either the firm or their clients would legally be held accountable.

These steps are largely performed manually, although the firm has experimented with automating various steps of the account-creation process. While past research on automated social bots has demonstrated the ways in which bots are used amplify certain content by gaming platform algorithms and piggybacking on strategic hashtags (Woolley, 2016; Murthy et al., 2016), the goal of these types of accounts is to persuade in a subtler manner. Outlining his firm's broader strategy, the research subject argued that their trolls/bots/influencers cannot, and do not attempt to influence public opinion directly. Rather, the firm's strategy is to target "opinion leaders", including journalists, politicians, bloggers, and key activists. By infiltrating influential Facebook groups, mining comment sections, and directly striking up conversations with these opinion leaders, the goal is to try to convince the target that their followers sincerely believe a certain argument and to provide long-term nudges towards certain strategically devised positions.

The amount of work which goes into these efforts is staggering, and the most involved campaigns will include multiple employees bringing their networks of accounts together to stage threads on discussion boards and steer conversations on forums. An entire thread on such a platform can feature dozens of fake accounts all posing as users, down-voting unsympathetic points of view, and generally steering a conversation in a form of what is often termed "astroturfing" (Woolley, 2016). All this occurs invisibly and behind the scenes, and the ordinary person that logs onto these forums may believe that they are receiving a legitimate signal for public opinion on a topic when they are in effect being fed a narrative by a secret marketing campaign.

While the current academic discussion predominantly focuses on automated bots, The Firm believes that their uses are limited because they are not able to interact with real users in a sophisticated manner. According to the research subject, political bots that try to directly impact discussion are highly inelegant and will almost certainly be discovered. Because a client must never be linked to these fake accounts, their company only uses truly automated bots for (a) spam and hate, and (b) as a red herring designed to discredit another actor. In the first case, the accounts used need not be highly sophisticated as they are not designed to persuade, but rather to spam and to perhaps influence platform algorithms (bots that retweet a negative story about a political figure, for example, can spread it widely by helping it "trend" on Twitter). In the second scenario, they would try to discredit another candidate by building network of obvious bots that would pose as that candidate's followers, spamming forums and harassing others in the name of another candidate, making it seem as if the rival candidate was employing bots and trolls.

A recent Facebook report, titled "Information Operations and Facebook," seems to corroborate the information provided by The Firm's employees. The paper, authored by members of Facebook's security team, provides the first public acknowledgement that state and non-state actors have been using a variety of "false amplifiers", such as fake accounts, bots, and astroturf groups filled with fake users, to influence political debate on the platform.

The authors suggest that Facebook's sophisticated anti-spam mechanisms are effective at thwarting most methods of automation, and instead, argue that Facebook is more concerned by manually controlled and created fake accounts

(Weedon et al., 2017). They note that much of this activity, such as the targeted infiltration of influential Facebook groups and pages, "could only be performed by people with language skills and a basic knowledge of the political situation in the target countries, suggesting a higher level of coordination and forethought" akin to that displayed by The Firm's employees. These types of manual influence efforts pose a particularly difficult problem for Facebook, as for privacy reasons it must find ways to find ways to flag fake accounts without directly screening content *en masse*. A new push on this front has yielded some success, with Facebook apparently removing some 30 000 fake accounts in the context of the 2017 presidential election in France (Weedon et al., 2017). While platforms are beginning to crack down on fake accounts, their prevalence on Polish social networks is likely to remain an issue in the foreseeable future.

Automated Accounts on Twitter

If fake accounts and false amplifiers exist on Polish Facebook, do they also exist on other platforms, such as Twitter? Despite being less frequently used by the general public, Twitter remains an important platform for Polish opinion leaders, politicians, and journalists. Several interviewees expressed their belief that automated bots do operate on Polish Twitter, albeit in small numbers, and many had anecdotal evidence that certain issues or hashtags had, in various cases, trended and spread in ways which seemed artificial or suspicious. However, research has yet to explore this possibility in a systematic way.

Methods and Limitations

In order to explore this issue further, a Twitter analysis was performed as follows. First, a list of 50 important political accounts was created. Using the Twittercounter service (which compiles a list of the 100 most followed users in a country), the most followed Polish political accounts were selected.[2] The list was rounded up to 50 relevant accounts using the author's best judgement, so that the final list included news organizations, politicians, journalists, and the official accounts of all major Polish parties. The full timelines of all 50 accounts were then downloaded using the Twitter Search API, and all tweets posted by these accounts over their lifetime were parsed for keywords and hashtags. This approach provided a snapshot of important political topics discussed over the last few years, and allows for the collection of

[2] *Twittercounter.com*

tweets with hashtags which have been consistently shown to be political and controversial on Polish twitter, including common hashtags such as #sejm (the Polish Parliament), as well as #smolensk (the site of the plane crash which killed a former Polish President, Lech Kaczyński), and #aborcja (abortion, another major political issue). This approach also mitigates at least some of the selection bias inherent in a research design where hashtags are selected by the researchers. A Streaming API query was set up using the top 30 political hashtags collected in this manner, and data was collected for three weeks in March and April 2017, yielding a dataset of 50 058 tweets.

The next step was to assess the level of automation within the sample dataset. Detecting bots on Twitter is not easy, and detecting political bots is even more difficult. Unlike other platforms which have large amounts of bot activity, such as Wikipedia, Twitter bot-makers are not required to register with some sort of central authority and overtly label their account as a bot. As Twitter has an incentive to underreport the number of bots on its platform, and also limits the data it will directly share with researchers, computer scientists have in recent years developed many complex models for detecting bots, with most systems implementing machine learning models based on certain account features, such as tweet frequency and social network characteristics (Ferrara et al., 2014). On the other hand, journalists and social science researchers interested in political bots generally use manual heuristics or simple thresholds to define automated activity. For example, our own project at the Oxford Internet Institute has in past defined bot accounts as accounts which tweet more than fifty times in a day on a certain hashtag (Howard & Kollanyi, 2016).

This approach, while providing a level of simplicity that is very useful when working with extremely large datasets (such as the one analyzed by our team in the lead up to the 2016 US election), is nevertheless limited by the method of data collection. Collecting tweets with certain hashtags means that this threshold will miss high-frequency accounts that only occasionally tweet using the queried hashtags, and instead tweet with other hashtags, reply directly to users, retweet content, or do not use hashtags at all. Therefore, this methodology may actually underestimate the amount of automated activity on political topics.

To help remedy this issue, four heuristics were utilized to better estimate bot activity in our dataset. First, the posting source was scrutinized: suspected bots can be

occasionally be identified by the various unusual ways that they use the Twitter API, which manifests itself as the 'posting source' metadata present with each downloaded tweet. For most tweets, this source is either the Twitter web client or an Android or iOS mobile client, or a social media management tool such as TweetDeck. But an analysis of the source metadata illustrates a long-tail of many custom sources: for example, an obvious Polish bot which claims to "share the most popular and top tweets of the day" (@Zaorany) uses a custom app titled "zaorane", and another account titled @haslaulicy uses an app titled "goodbot".

Second, a simple engagement ratio (tweets + favourites / number of followers) was created, with the logic being that if a user is posting and liking more than a thousand times for every follower they gain, there is a high probability of the account being a bot. This method flagged several obvious bots with a ratio of several thousand tweets for every follower.

The third technique entails looking at lifetime tweets: this is the total number of tweets an account has posted at the last point in the data collection divided by account's "age" (the number of days since its creation). This method flags accounts which may only have had a few dozen tweets captured in our dataset, but still are tweeting very actively on other hashtags or are not using hashtags.

The final heuristic involves assessing the total number of tweets posted by an account in the data collection period. For example, if an account has several tweets in our dataset, including one on the first day of collection and one three weeks later, the changes in the "total number of account tweets" metadata provided by the Twitter API can be compared, illustrating the overall number of tweets posted in that period. While it is generally not possible to state with perfect certainty that an account is a bot unless it self identifies as such, these four heuristics provide a way to flag suspicious accounts for further investigation.

Analysis

Through a combination of these four heuristics, a short list of 500 suspect accounts was created. These 500 accounts (which form 4.97% of the total 10 050 unique accounts in the dataset) tend to tweet at a much higher frequency than the average user, and were responsible for a disproportionately large amount of tweets in the dataset, generating 16937 observed tweets (33.8% of the collected sample of 50 058 tweets). These accounts were then evaluated and manually coded into a

grounded typology with four main categories. The bio, profile photo, cover photo, and fifteen most recent tweets of each account were assessed and categorized in a content analysis as follows:

- Right Wing

This category was composed of accounts which openly support the current PiS government, as well as more extreme right-wing nationalist accounts. Accounts in the former category tended to prominently self-identify with PiS, either through their username (eg. @rutkowski1PiS), or their bio, in which they would openly state that they were conservatives and PiS supporters. Nationalist ideology and slogans often collided with party language, making these accounts very easy to identify (for example, @Kriskrak197 loudly proclaims *BÓG HONOR, OJCZYZNA*", translated as "God, honor, and the fatherland", below a cover photo featuring a word cloud with the official pro-PiS hashtags).[3] These accounts were further identified by their tendency to propagate hyper-partisan content, especially around the Smolensk disaster and other sensitive political issues, and retweet content from a network of accounts that question the veracity of left-leaning media outlets, such as *TVN* and *Gazeta Wyborcza*

. Extreme accounts tended to be even more aggressive in their nationalist rhetoric, claiming for example that "leftism is the cancer which is devouring Poland" ("*Lewactwo to jest rak który zżera Polskę*") or that left-leaning individuals were criminally insane or subhuman (*Appendix D).* The content retweeted was often inflammatory and featured a very strong stance against immigration, Muslims, the European Union, Poland's Civic Platform party, and the Committee for the Defense of Democracy, a Polish NGO that opposes the current government. Prominent themes observed among these more nationalist accounts included the framing of immigration in Europe as a holy war akin to the Crusades, skepticism about the investigations into the Smolensk disaster, and, interestingly, pro-gun and anti-feminist content linked to the US alt-right movement via influencers such as Jack Posobiec and Mike Cernovich. The main news sources retweeted in the right-wing camp included @TVP, @RepublikaTV, and @wPolityce_pl, as well as the noted alt-right account @V_of_Europe, which tweets questionable stories from Breitbart, Infowars, and RT.

[3] See appendix C.

- Left Wing

This category was composed of accounts which clearly identified with the PO (the centrist opposition party), KOD, Razem (a left-wing party that does not currently have seats in the Polish Sejm), or overtly expressed their opposition to the ruling PiS government. These accounts tended to feature bios with hashtags such as #*NieDlaPiS* ("not for PiS"), #*StopPis*, and #*Opozycya* ("Opposition"), and tended to retweet stories from media organizations such as TVN and Gazeta Wyborcza, influential politicians and journalists such as Bartosz Wielinski, Tomasz Siemoniak, Borys Budka, and ideologically aligned Twitter influencers such as @Marcin_Kaminski, @lis_tomasz, and the meme-heavy @SOKzBURAKApl. Accounts apparently aligned with a new account titled the "Citizen's Opposition" (@OObywatelska) were also coded in this category (including a number of suspect accounts displaying a profile picture with the "Citizen's Opposition" logo in the bottom right corner).

- Neutral

This category included accounts that shared content from a combination of right and left wing media outlets (such as @Tylko_newsy). If the political leaning of an account was not immediately obvious, it was classified as neutral.[4]

- Other

This category included a variety of accounts which did not fit into the above three categories. Accounts which were no longer accessible at time of coding, due to having been suspended or removed, were coded as "Inaccessible". Accounts which seemed non-political and were sharing spam, marketing, or pornographic content were coded as "Junk/Spam/Porn". Suspect accounts which turned out to be verified news outlets using content management software were coded as "Verified". These accounts included Rzeczpospolita, Sputnik Polska, and ONET News. Finally, accounts which had a bio in Ukrainian or Russian, and tweeted predominantly Ukrainian or Russian content, were coded at "Ukraine" and "Russia", respectively.

The preliminary results of this analysis show that in the sample collected, there are more than twice as many suspicious right-wing accounts as there are left-wing accounts. These accounts are highly prolific: the 263 right-wing accounts that were

[4] Note: this data was coded by the author, a left-leaning individual who receives his news primarily through mainstream outlets. This may have affected how the content shared by these accounts was perceived and categorized.

coded here generated 10 053 tweets in the sampled dataset. In comparison, the 113 suspect accounts coded as left-wing only generated 2073 tweets. The 263 suspect right-wing accounts were responsible for 20.0% of all tweets in the dataset collected.

Further research will be required to unpack and map this ecosystem of right-wing automation, as many of these accounts appeared to not only be sharing content from certain influencers, but actively retweeting other obvious right-wing bots coded in this analysis. The accounts perpetuate fringe points of view and spread political disinformation from untrustworthy channels, such as Voice of Europe, or partisan blogs posing as legitimate news outlets. For example, the right-wing accounts were observed widely sharing content from @wPrawopl, an account created in May that claims to be a news portal, but on closer inspection seems to be the personal project of a Polish YouTuber that spreads inflammatory stories and conspiracy theories such as "How the Jews helped the Germans murder the Jews" with a stated mission of "teaching Poles the truth".

Figure 1: Classification of Suspected Bot Accounts

Type of Account	*N*	*%*
Right Wing	263	52.6
Left Wing	113	22.6
Neutral	45	9.0
Other:		
Inaccessible	28	5.6
Junk/Spam/Porn	22	4.4
Verified	14	2.8
Russia	8	1.6
Ukraine	7	1.4
Total	500	100

Source: Author's calculations from data sampled 21 March – 12 April 2017, and coded on the 8-10th of June. Hashtags include: #wolnemediawsejmie, #ukraina, #terazwsejmie, #szczytnato, #stanwojenny, #solidarność, #smoleńsk, #sejmprotest, #sejmie, #sejm, #samoobronakobiet, #rosja, #polskiemedia, #planrozwojupl, #pamiętamy, #obronaterytorialna, #morawiecki, #litwa, #katastrofasmoleńska, #kaczyński, #gruzja, #funduszeeuropejskie, #dzieńbezpolityków, #dezinformacja, #czarnyprotest, #bezpieczeństwo, #ambergold, #aborcja

One of the disturbing aspects of this is that apparent bots are spreading highly inflammatory and oftentimes xenophobic content which may be seen by ordinary users (*Appendix E*), a particularly problematic development if social media help users formulate signals for public opinion. Bots can coordinate action on strategic hashtags to generate the appearance of public support, a process which Woolley has called "manufacturing consensus" (Manjoo, 2017). In the case of Poland, this could be true on a number of levels, with posts from the official PiS accounts actively being retweeted and liked by a number of automated accounts, and

therefore appearing more popular when observed by journalists, other politicians, and ordinary users. This same process could also influence more insular communities, such as Polish right-wing nationalist groups on Twitter and the individuals who sympathize which their viewpoints, who could be convinced that the content they are engaging with is more influential than it actually is, resulting in a sort of algorithmic, bot-driven confirmation bias that could have negative long-term effects.

Overall, the preliminary results of this analysis suggest that there are higher levels of automation on Polish twitter than previously thought. The total percentage of traffic that can be attributed to all suspect accounts (33.8%), is as almost as high as the share of automated traffic on pro-Trump hashtags (known to feature high levels of bot activity) in the lead up to the 2016 Presidential debates (35.9%; see Howard & Kollanyi, 2016). It also seems that the everyday share of automated content on Polish Twitter (as the data collection was not timed to coincide with any specifically contentious or important political events) is considerably higher than what has been observed in the lead up to elections in France, Germany, and the United Kingdom (Gallacher et al., 2017).

Additionally, the finding that right-wing pro-government and nationalist accounts form the majority of suspect bot accounts, and indeed are far more prolific on the collected political hashtags than their left-wing counterparts, provides another element of evidence to corroborate the commonly held assumption that the Polish right has been more effective online, having implemented a variety of new tools and practices more effectively than their rivals. Finally, the analysis demonstrates non-insignificant numbers of apparent left wing "opposition" bots, many of which seem to have been created in March of this year, and could be part of a concerted campaign to battle back against the perceived influence of the Polish right-wing digital ecosystem. These developments will need to be mapped and further assessed over the coming months.

Conclusion

The Internet's architecture and affordances of anonymity not only make it very difficult to impede the various mechanisms of computational propaganda, but also to simply get an understanding of their scope and scale. From detailed efforts to influence via meticulously crafted fake accounts on Facebook, to networks of

automated Twitter accounts that attempt to megaphone content, if these sorts of practices are happening in Poland, then it seems especially likely that they are happening in other countries. But how prevalent is this activity, really? And what kind of effects does it really have on political discourse?

First of all, one needs to reflect upon the problem of fake accounts on Facebook. As the company interviewed has been engaging in this type of activity for over ten years, it is likely that networks of artificial identities have been deployed on Facebook by other actors for a long period of time as well. (Indeed, it is likely that the various false amplifiers discussed in this report have existed for a while, but have only become the focus of mainstream public debate and discussion in the West since the 2016 US election). These practices, as described by The Firm's employees and Facebook's security team, pose several questions and challenges for researchers. The first type of challenge is a theoretical one. These accounts are not quite bots, but not quite trolls as traditionally conceived in the online political communication literature either. In many ways, they blur the lines between political marketing and propaganda, as the same techniques could in effect be transitioned seamlessly from the commercial space (to benefit a firm or industry) to the political space (to benefit a party or candidate). The second set of challenges features various methodological problems. How should academics best study these false amplifiers, which have been confirmed by Facebook itself as having an important influence on political debate, but operate invisibly and behind the scenes on closed platforms that withhold data from researchers? Without concrete data, it becomes very difficult to measure the true scope and scale of these efforts, and to empirically determine their actual effects on users.

Secondly, Twitter bots need to be better understood. While we know that they can have an amplifying effect on content and help game trending algorithms, to what extent do they really affect the experience of the average user, especially if they are simply engaging with content created within their potentially insular groups of friends and followers? How much do they really influence political opinions over time? What role exactly do these accounts play within the larger disinformation ecosystem, and how exactly do they coordinate to potentially spread hyper-partisan "fake news"?

These are increasingly important questions, as we rapidly seem to be entering a new golden age of propaganda, misinformation, and media manipulation,

compounded by the wide-ranging political instability and electoral uncertainty that has characterized European politics of late. We must better understand these developments before we can truly begin to craft solutions.

A look at Poland provides insight into the complexities of studying computational propaganda today, and provides some new perspectives into what is rapidly becoming a global phenomenon. Overall, the findings suggest that false amplifiers are indeed prevalent on both Polish Facebook and Twitter, and that further research should be conducted in this area.

References

Allcott, H., & Gentzkow, M. (2017). *Social Media and Fake News in the 2016 Election* (National Bureau of Economic Research Working Paper). Rochester, NY. Retrieved from https://papers.ssrn.com/abstract=2903810

Baranowski, P. (2015). Online Political Campaigning during the 2014 Regional Elections in Poland. *Media and Communication*, *3*(4), 35–44.

Batorski. (2015). Social filtering on the Internet – a new mechanism of content curation and its consequences. *Media Studies*, *62*(3). Retrieved from http://studiamedioznawcze.pl/article.php?date=2015_3_62&content=batorski&lang=en

Batorski, D., & Grzywińska, I. (2017). Three dimensions of the public sphere on Facebook. *Information, Communication & Society*, *0*(0), 1–19.

Beyer, J. (2014). *Expect Us: Online Communities and Political Mobilization*. New York: Oxford University Press.

CERT Polska (2015). "Krajobraz bezpieczeństwa polskiego internetu." Retrieved from https://goo.gl/4MI0p3

Chen, A. (2015, June 2). The Agency. *The New York Times*. Retrieved from https://www.nytimes.com/2015/06/07/magazine/the-agency.html

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on Twitter: human, bot, or cyborg? In *Proceedings of the 26th annual computer security applications conference* (pp. 21–30). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1920265

Coleman, E. G. (2012). Phreaks, Hackers, And Trolls: The Politics Of Transgression And Spectacle. In Mandiberg, Michael (Ed.), *The Social Media Reader*. New York: New York University Press.

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* (pp. 273–274). https://doi.org/10.1145/2872518.2889302

Dubiński, P. (2015, October 23). Internet znów przesądzi o wyniku wyborów? Eksperci nie mają wątpliwości. Retrieved from http://wiadomosci.wp.pl/internet-znow-przesadzi-o-wyniku-wyborow-eksperci-nie-maja-watpliwosci-6027738241049217a

Eurobarometer (2016). Internet access and use statistics – households and individuals. Retrieved from http://ec.europa.eu/eurostat/statistics-explained/index.php/Internet_access_and_use_statistics_-_households_and_individuals#Database

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The Rise of Social Bots. *Communications of the ACM*, *59*(7), 96–104.

Gallacher, J.D., Kaminska, M., Kollanyi, B., Yasseri, T., & Howard, P.N. Social Media and News Sources during the 2017 UK General Election." Data Memo 2017.6. Oxford, UK: Project on Computational Propaganda. comprop.oii.ox.ac.uk.

Gemius/PBI (2017). Poland Internet Statistics, Retreived from http://www.wirtualnemedia.pl/artykul/wyniki-badania-gemius-pbi-za-luty-2017

Głowacka, D., Ploszka, A., & Sczaniecki, M. (2016). Wiem i powiem: Ochrona sygnalistów i dziennikarskich źródeł informacji. Warsaw, Poland: Helskinki Foundation for Human Rights.

Głowacki, W. (2015, September 28). Prawo i Sprawiedliwość króluje w polskim internecie. Pomaga w tym zdyscyplinowana armia trolli. *Gazeta Krakowska*. Retrieved from http://www.gazetakrakowska.pl/artykul/8866523,prawo-i-sprawiedliwosc-kroluje-w-polskim-internecie-pomaga-w-tym-zdyscyplinowana-armia-trolli,id,t.html

Guilbeault, D. (2016). Growing Bot Security: An Ecological View of Bot Agency. *International Journal of Communication*, *10*(0), 19.

Herring, S., Job-Sluder, K., Scheckler, R., & Barab, S. (2002). Searching for Safety Online: Managing "Trolling" in a Feminist Forum. *The Information Society*, *18*(5), 371–384.

Howard, P. N., & Kollanyi, B. (2016). *Bots, #Strongerin, and #Brexit: Computational Propaganda during the UK-EU Referendum* (Working Paper No. 2016.1). Oxford, UK: Project on Computational Propaganda. Retrieved from www.comprop.oii.ox.ac.uk

Koc-Michalska, K., Lilleker, D. G., Smith, A., & Weissmann, D. (2016). The normalization of online campaigning in the web.2.0 era. *European Journal of Communication*, *31*(3), 331–350.

Koc-Michalska, K., Lilleker, D. G., Surowiec, P., & Baranowski, P. (2014). Poland's 2011 Online Election Campaign: New Tools, New Professionalism, New Ways to Win Votes. *Journal of Information Technology & Politics*, *11*(2), 186–205.

Kollanyi, B. (2016). Where Do Bots Come From? An Analysis of Bot Codes Shared on GitHub. *International Journal of Communication*, *10*, 20.

Kollanyi, B., Howard, P. N., & Woolley, S. C. (2016). *Bots and Automation over Twitter during the U.S. Election* (Data Memo No. 2016.4) (p. 5). Oxford, UK: Project on Computational Propaganda. Retrieved from comprop.oii.ox.ac.uk

Krzymowski, M. (2016, June 5). Ucho partii. *Newsweek Polska*. Retrieved May 30, 2017, from http://www.newsweek.pl/plus/polska/pawel-szefernaker-kim-jest-internetowy-geniusz-pis-,artykuly,386767,1,z.html

Kublik, A. (2015, January 2). Rząd bierze media publiczne. Retrieved May 30, 2017, from http://wyborcza.pl/1,75398,19419297,rzad-bierze-media-publiczne.html?disableRedirects=true

Lee, K., Eoff, B. D., & Caverlee, J. (2011). Seven months with the devils: a long-term study of content polluters on Twitter. In *AAAI Int'l Conference on Weblogs and Social Media (ICWSM)*.

Lucas, E., & Nimmo, B. (2016). *Winning the Information War: Techniques and Counter-strategies to Russian Propaganda in Central and Eastern Europe*.

Centre For European Policy Analysis. Retrieved from http://cepa.org/reports/winning-the-Information-War

Maréchal, N. (2016). When Bots Tweet: Toward a Normative Framework for Bots on Social Networking Sites. *International Journal of Communication* 10, 10.

Manjoo, F. (2017, May 31). How Twitter Is Being Gamed to Feed Misinformation. *The New York Times*. Retrieved from https://www.nytimes.com/2017/05/31/technology/how-twitter-is-being-gamed-to-feed-misinformation.html

Marwick, A., & Lewis, R. (2017). Media Manipulation and Disinformation Online. Data & Society Research Institute Report. https://datasociety.net/output/media-manipulation-and-disinfo-online/

Mitter, S., Wagner, C., & Strohmaier, M. (2014). A categorization scheme for socialbot attacks in online social networks. *arXiv:1402.6288 [physics]*. Retrieved from http://arxiv.org/abs/1402.6288

Murthy, D., Powell, A. B., Tinati, R., Anstead, N., Carr, L., Halford, S. J., & Weal, M. (2016). Bots and Political Influence: A Sociotechnical Investigation of Social Network Capital. *International Journal of Communication* 10, 20. Retrieved from http://ijoc.org/index.php/ijoc/article/view/6271

Napiórkowski, M. (2017, March 21). Niebieski wieloryb. List z Ministerstwa Edukacji Narodowej. Retrieved May 30, 2017, from http://mitologiawspolczesna.pl/niebieski-wieloryb-list-ministerstwa-edukacji-narodowej/

Olwert, P., & Wachnicki, M. (2014, March 4). Wynajęci Rosjanie cyber-bombardują polski internet? *Newsweek Polska*. Retrieved February 14, 2017, from http://www.newsweek.pl/swiat/wynajeci-rosjanie-cyber-bombarduja-polski-internet-newsweek-cyberatak,artykuly,281538,1.html

Ostrowki, W., & Woycicki, K. (2016). Case Study: Poland. In *Winning the Information War: Techniques and Counter-strategies to Russian Propaganda in Central and Eastern Europe*. Centre For European Policy Analysis. Retrieved from http://cepa.org/reports/winning-the-Information-War

Pfetsch, B., & Voltmer, K. (2012). Negotiating Control. *The International Journal of Press/Politics*, *17*(4), 388–406. https://doi.org/10.1177/1940161212449084

Pomerantsev, P., & Weiss, M. (2014). The Menace of Unreality: How the Kremlin Weaponizes Information, Culture and Money. *The Interpreter*. Retrieved from http://www.interpretermag.com/the-menace-of-unreality-how-the-kremlin-weaponizes-information-culture-and-money/

Przełomiec, M. (2014). Poland on the Euromaidan. In Bachmann, Klaus & Lyubashenko, Igor (Eds.), *The Maidan Uprising, Separatism and Foreign Intervention: Ukraine's Complex Transition* (pp. 299–314). Frankfurt: Peter Lang.

Rankin, J., & Traynor, I. (2016, January 12). European commission to debate Poland's controversial new laws. *The Guardian*. Retrieved from

https://www.theguardian.com/world/2016/jan/12/european-commission-to-debate-polands-controversial-new-laws

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Detecting and Tracking the Spread of Astroturf Memes in Microblog Streams. *arXiv:1011.3768 [cs]*, 249. https://doi.org/10.1145/1963192.1963301

Rid, T., & Buchanan, B. (2015). Attributing Cyber Attacks. *Journal of Strategic Studies*, *38*(1-2), 4–37. https://doi.org/10.1080/01402390.2014.977382

Savytskyi, Y. (2016, June 20). Kremlin trolls are engaged in massive anti-Ukrainian propaganda in Poland. *Euromaidan Press*. Retrieved December 15, 2016, from http://euromaidanpress.com/2016/06/21/kremlin-trolls-are-engaged-in-massive-anti-ukrainian-propaganda-in-poland/

Simons, T. W. (2008). *Eurasia's new frontiers: young states, old societies, open futures*. Ithaca, NY: Cornell University Press.

Sobkowicz, P., & Sobkowicz, A. (2012). Two-Year Study of Emotion and Communication Patterns in a Highly Polarized Political Discussion Forum. *Social Science Computer Review*, *30*(4), 448–469.

Sotrender. (2016a, January 28). Facebook w Polsce - podsumowanie 2015. Retrieved May 30, 2017, from https://www.sotrender.com/blog/pl/2016/01/facebook-w-polsce-podsumowanie-2015-r-infografika/

Sotrender. (2016b, January 27). Twitter w Polsce - podsumowanie. Retrieved May 30, 2017, from https://www.sotrender.com/blog/pl/2016/01/twitter-w-polsce-podsumowanie-2015-r-infografika/

Sputnik Polska. (2017, February 20). Jak rozpoznać rosyjskiego trolla? Retrieved February 23, 2017, from https://pl.sputniknews.com/polityka/201702204869717-Sputnik-Rosja-trolling/

Starbird, K. (2017). Examining the Alternative Media Ecosystem through the Production of Alternative Narratives of Mass Shooting Events on Twitter. In *11th International AAAI Conference on Web and Social Media (ICWSM)*.

Sunstein, C. R., & Vermeule, A. (2008). *Conspiracy Theories* (SSRN Scholarly Paper No. ID 1084585). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=1084585

Szczepaniak, P., & Szczygieł, K. (2017, March 5). Polskie fejki, rosyjska dezinformacja. OKO.press tropi tych, którzy je produkują. Niektórzy z nich nie istnieją. *OKO Press*. Retrieved from https://oko.press/wszystkie-media-popelniaja-bledy-niektore-robia-celowo/

Tismaneanu, V. (2009). *Fantasies of salvation: Democracy, nationalism, and myth in post-communist Europe*. Princeton, NJ: Princeton University Press.

Trammell, K. D., Tarkowski, A., Hofmokl, J., & Sapp, A. M. (2006). Rzeczpospolita blogów [Republic of Blog]: Examining Polish Bloggers Through Content Analysis. *Journal of Computer-Mediated Communication*, *11*(3), 702–722.

Tsvetkova, M., García-Gavilanes, R., Floridi, L., & Yasseri, T. (2017). Even good bots fight: The case of Wikipedia. *PLOS ONE*, *12*(2). https://doi.org/10.1371/journal.pone.0171774

Urbanek, G. (2016, November 3). Facebook odblokowuje konta. Narodowcy nie składają broni - Kraj. *Rzeczpospolita*. Retrieved from http://www.rp.pl/Kraj/311039855-Facebook-odblokowuje-konta-Narodowcy-nie-skladaja-broni.html

Weedon, J., Nuland, W., & Stamos, A. (2017). *Information Operations and Facebook*. Facebook Security. Retreived from https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf

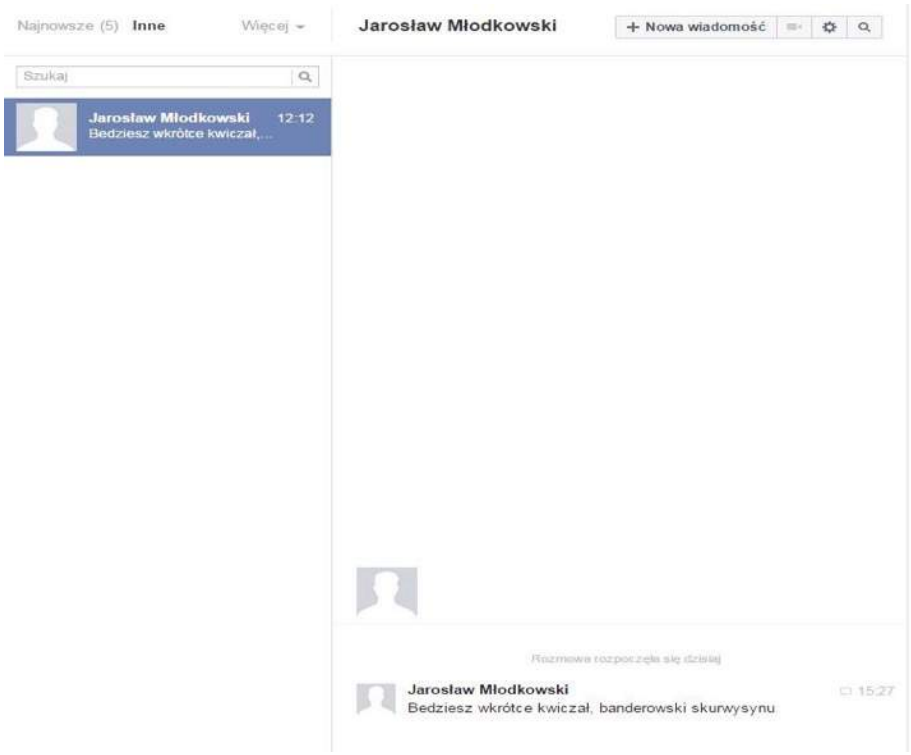
Wieliński, B. (2015, December 11). Czerwone brygady PiS w internecie. *Gateza Wyborcza*. Retrieved from http://wyborcza.pl/1,75968,19331666,czerwone-brygady-pis-w-internecie.html

Wierzejski, A. (2016). *Information Warfare In The Internet: Exposing and Countering Pro-Kremlin Disinformation in the CEEC*. Centre For International Relations. Retrieved from http://csm.org.pl/en/publications

Woolley, S. C. (2016). Automating power: Social bot interference in global politics. *First Monday*, *21*(4).

Woolley, S. C., & Howard, P. N. (2016). Political Communication, Computational Propaganda, and Autonomous Agents—Introduction. *International Journal of Communication* 10, 9.

Appendix A: Sample Threat Delivered to Journalist[5]



Appendix B: AntiFa Proclaims Victory

[5] Source for Appendix A: Research subject 'Daydream'. Appendix B: Author's own screenshot, Oct. 27, 2016. Appendix C-G: Author's own screenshots, June 12, 2017.

Appendix C: Example Suspect Account

Appendix D: Example Suspect Account

Appendix E: Example images retweeted by suspect accounts

Appendix F: Example Suspect Account

About the Author

Robert Gorwa is a graduate student at the Oxford Internet Institute, University of Oxford. He studies the effects of technology on various international relations phenomena, with a focus on the internet and its implications for international security, inter- and intra-state relations, and conflict. He begins his PhD in Oxford's Department of Politics and International Relations in the fall of 2017.

Citation

Robert Gorwa, "Computational Propaganda in Poland: False Amplifiers and the Digital Public Sphere." Samuel Woolley and Philip N. Howard, Eds. Working Paper 2017.2. Oxford, UK: Project on Computational Propaganda. comprop.oii.ox.ac.uk<http://comprop.oii.ox.ac.uk/>. 32 pp.

Series Acknowledgements

The authors gratefully acknowledge the support of the European Research Council, Computational Propaganda: Investigating the Impact of Algorithms and Bots on Political Discourse in Europe," Proposal 648311, 2015-2020, Philip N. Howard, Principal Investigator. Additional support has been provided by the Ford Foundation and Google-Jigsaw. Project activities were approved by the University of Oxford's Research Ethics Committee. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders or the University.