EAGER: Computational Propaganda and The Production / Detection of Bots

NSF Org:	SES Division of Social and Economic Sciences
Initial Amendment Date:	August 1, 2014
Latest Amendment Date:	August 1, 2014
Award Number:	1450193
Award Instrument:	Standard Grant
Program Manager:	Heng Xu SES Division of Social and Economic Sciences SBE Directorate for Social, Behavioral & Economic Sciences
Start Date:	September 15, 2014
Expires:	August 31, 2016 (Estimated)
Awarded Amount to Date:	$218,825.00
Investigator(s):	Philip Howard pnhoward@u.washington.edu (Principal Investigator) David McDonald (Co-Principal Investigator)
Sponsor:	University of Washington 4333 Brooklyn Ave NE SEATTLE, WA 98195-9472 (206)543-4043
NSF Program(s):	Secure &Trustworthy Cyberspace
Program Reference Code(s):	7434, 7916, 8225
Program Element Code(s):	8060

ABSTRACT

Political bots are manipulating public opinion over major social networking applications. This project enables a new team of social and information scientists to investigate the impact of automated scripts, commonly called bots, on social media. The PIs will study both the bot scripts and the people making such bots, and then work with computer scientists to improve the way we catch and stop such bots. Experience suggests that political bots are most likely to appear during an international crisis, and are usually designed to promote the interests of a government in trouble. Political actors have used bots to manipulate conversations, demobilize opposition, and generate false support on popular sites like Twitter and Facebook from the U.S. as well as Sina Weibo from China.

The first stage of this research is international fieldwork with the political consultants and computer experts who are commissioned to make bots. Second, the PIs are building an original database of political incidents involving bots. Finally, the PIs are using this knowledge to make better tools for detecting political bots when they appear. The PIs are doing "real-time" social and information science, and actively disseminating their findings to journalists, industry, and foreign policy experts. By developing an a network of experts in political bot detection and an original data set, the researchers will not only have a better understanding of how bots are manipulating social networks but also advance the conversation in the social sciences, computer sciences, and industry about the size of the problem and the possible solutions.

PROJECT DESCRIPTION

1. Introduction

Social media has revolutionized the way people discuss current affairs and obtain political news and information. Contact with social media helps young people cultivate a political identity and engage civically in both authoritarian and democratic regimes (Bennett and Segerberg 2013, Howard 2011). Activist causes and democratic movements have been born, organized and disseminated on sites like Facebook, Twitter, Weibo, and YouTube. The positive, and wide-reaching, democratic potential of social media is much discussed, but another, more propagandistic, side of this new technology exists (Diamond 2010, Shirky 2008, 2010, Woolley and Howard, forthcoming). Security experts argue that more than 10 percent of content on social media websites, and 62 percent of all web traffic, is generated by bots (Rosenberg, 2013). This EAGER project will assemble a new team of multidisciplinary researchers to work on a three-stage research agenda: a) international fieldwork (participant observation, interviews) with teams of bot makers and the people building bot detectors; b) construction of an original event-dataset of known incidents of bot use by political actors and c) computational theory about how bots can be detected during sensitive moments when the manipulation of public opinion would have detrimental effects on public life.

The word "botnet" comes from combining "robot" with "network." It is used to describe a collection of programs that communicate across multiple devices to perform some task. The tasks can be simple and annoying, like generating spam. The tasks can be aggressive and malicious, like choking off exchange points, or launching denial-of-service attacks. And not all are developed to advance political causes. Some seem to have been developed for fun or to support criminal enterprises, but all share the property of deploying messages and replicating themselves (Kim el al. 2010, Wagstaff, 2013). Chu et al. distinguish two types of bots on Twitter: legitimate and malicious. Legitimate bots generate a large amount of benign tweets that deliver news or update feeds. Malicious bots, on the other hand, spread spam by delivering appealing text content with the link directed to spam or malicious content. Botnets are created for many reasons: spam, DDoS attacks, theft of confidential information, click fraud, cyber sabotage, and cyber warfare. According to Kim et al., many governments have been strengthening their cyber warfare capabilities for both defensive and offensive purposes. In addition, political actors and governments worldwide have begun using bots to manipulate public opinion, choke off debate, and muddy political issues.

Social bots are particularly prevalent on Twitter. They are computer generated programs that post, tweet, or message of their own accord. Often bot profiles lack basic account information such as screen-names or profile pictures. Such accounts have become known as "Twitter eggs" because default profile pictures on the social media site ubiquitously feature a egg. While social media users access from front-end websites, bots get access to such websites directly through a mainline, code-to-code, connection, mainly, through the site's wide-open application programming interface (API), posting and parsing information in real time. Bots are versatile, cheap to produce, and ever evolving. "These bots," argues Rob Dubbin, "whose DNA can be written in almost and modern programming language, live on cloud servers, which never go dark and grow cheaper by day (Dubbin, 2013)." Over the last two decades, with a rapid increase in just the last two years, developers and their employers have begun to deploy bots beyond mundane commercial tasks like spamming or scraping sites like eBay for bargains. Bots are the primary applications used in carrying out distributed denial of service and virus attacks, email harvesting, and website content theft. Beyond this, and central to our research, lies the explicitly political usage of bots as social media propaganda tools.

The use of political bots varies across regime types. As a preliminary exercise we hope to construct a comprehensive typology of worldwide political bot-usage. Our current understandings, based upon initial research, suggest that political bots tend to be used for distinct purposes during three primary events: elections, spin control during political scandals, and national security crises. The usage of bots during these situations extends from the nefarious cause of demobilizing political opposition followers to the *seemingly* innocuous task of padding political candidates' social media "follower" lists. Bots are additionally used to drown-out oppositional or marginal voices, halt protest, and relay "astroturf" messages of false governmental support. Political actors use them in general attempts to manipulate and sway public opinion. It is clear that understanding the creation and usage of this technology is central to generating political equality both on and off line and in fostering genuine advancement of democratic social media possibilities. We believe that our

collaborative project will greatly illuminate several aspects of political bot creation, usage, and effect for communities professional, public, and policy oriented.

1.1. The Proliferation of Bots and Computational Propaganda

What triggered this EAGER proposal was a shared interest in the computational propaganda being carried out by what we suspect to be programmers employed by The Syrian Electronic Army (SEA), a hacker network that supports the Syrian government. The group developed a botnet that generates pro-regime content with the aim of flooding the Syrian revolution hashtags (e.g. #Syria, #Hama, #Daraa ...etc.) and overwhelming the pro-revolution discussion on Twitter and other social media portals (Qtiesh, 2011, York, 2011). As the Syrian blogger Anas Qtiesh writes, "These accounts were believed to be manned by Syrian Mokhabarat (intelligence) agents with poor command of both written Arabic and English, and an endless arsenal of bite and insults (Guardian, 2011)."

Differing forms of bot generated computational propaganda have been deployed in several other countries: Russia, Mexico, China, Australia, the United Kingdom, the United States, Azerbaijan, Iran, Bahrain, South Korea, and Morocco. Current contemporary political crises in the Thailand, Turkey, and the ongoing situation in Ukraine are seeing the emergence of computational propaganda. Table 1 presents a casual sampling of the diversity of regime types and bot producers around the world, with a democracy score from -10 fully authoritarian to +10 fully democratic (Marshall et al, 2011). This preliminary case list suggests that bot usage is often associated with either elections or national security crises. These may be the two most sensitive moments for political actors where the potential stigma of being caught manipulating public opinion is not as serious as the threat of having public opinion turn the wrong way.

Most of the coverage of political bot usage has occurred within mainstream media sources and personal blogs. Little empirical social or computer science work has been done to understand the wide-ranging creation, use, and effect of computational propaganda. Existing research on the topic of bots is limited to studies developing rudimentary bot detection systems, how bots challenge network security, and overviews of bots and botnets—networks composed of bots. Current research fails to develop an understanding of the new political bot phenomena, does not adequately explain the usage of these bots on social media sites, and lacks in any attempt to understand the makers of this technology. Social and computer science research on astroturf campaigns, public opinion manipulation and computational propaganda has been occurring very independently from each other. Social scientists have tended to focus on the organizational culture of engineers and technologists that produce malware, and computer scientists have tended to focus on bot dissemination. Our team will work together to study both processes and contextualize political bots. While botnets have been actively tracked for several years, their use in political campaigning, crisis management and counter-insurgency is relatively new (Kim et al, 2010). Moreover, from the users' perspective it is increasingly difficult to distinguish between content that is generated by a fully automated script, a human, or both (Chu et al, 2010).

1.2. Bots and the Internet Census

The first "internet census" was conducted in 2012 by an unknown party. It is not clear that it was a scholarly endeavor but is accepted as a credible study of global botnets. She wrote code that would both count devices and replicate itself so that its copies could help count devices. When she activated the bot, it created a botnet that identified 1.3 billion IP addresses being used by devices around the world (Unknown, 2012). The author called her script the Carna Bot after the Roman goddess of health and vitality. She really did think the exercise was about taking

Table 1: Selected Incidents of Political Bot Usage, by Country

Country	Year	Polity IV Score	Suspected Deployer	Source
Australia	2013	10	State	(Peel, 2013)
Azerbaijan	2012	-8	State	(Pearce, 2013)
Bahrain	2011	-8	State, Outsourced	(York, 2011)
China	2012	-8	State	(Krebs, 2012)
Iran	2011	-6	State, Outsourced	(York, 2011)
Israel	2012	10	State	(Painter, 2013)
Mexico	2011	8	Political Parties	(Herrera, 2012)
Morocco	2011	-6	State, Outsourced	(York, 2011)
Russia	2011	4	State	(Krebs, 2011)
Saudi Arabia	2013	-10	State	(Freedom House, 2013)
South Korea	2012	8	State	(Sang-Hung, 2013)
Syria	2011	-8	State, Outsourced	(York, 2011)
Tibet	2012	-8	State	(Krebs, 2012)
UK	2012	10	State	(Downes, 2012)
US	2011	10	State, Outsource	(Coldeway, 2012)
Venezuela	2012	2	State	(Shields, 2013)

basic measurements of the health of the internet. Her bot worked brilliantly, reporting back basic information on many different kinds of devices, from web cams and consumer routers, to printers and door-security systems.

The author of the bot decided to remain anonymous but published her findings as a public service. Even though she had noble goals, she publicized two concerning trends with the social application of computing systems. First, she revealed that knowing the default passwords for four pieces of key equipment could give someone access to hundreds of thousands of consumer devices and tens of thousands of industrial devices around the world, from gaming platforms to industrial-control systems. So the world's security experts may be debating the impact of the latest complex hacking attempts from China or the encryption possibilities of quantum computers. Knowing the factory passwords means access to devices once they leave the factory and get connected to the internet.

Second and more concerning, the bot discovered another bot. Carna wasn't the only unauthorized bot checking for open ports on devices around the globe. The bot was written as a public service for an exploratory project, and it built a botnet to do the census. But she found several competing botnets, and an enormous, largely sleeping network of bots called the "Aidra botnet" that had compromised as many as 30,000 devices. The bot was designed to hijack not just computers, but gas meters, refrigerators, microwaves, car-management systems and some mobile phones. The bots could attack any network infrastructure for a client with a denial-of-service attack. The author had her Carna Bot perform the public service of temporarily disabling any Aidra bots they found.

But the next time someone reboots those infected devices, the bots will be ready to start commandeering devices. Obviously there is a lot of destructive potential behind the malicious botnet exposed, and some might even see her as a threat because she also wrote a script that interfered with network traffic and device function. Understanding the production, dissemination and use of bots and botnets requires tracing the process of engineering decisions, innovations by computer scientists and hackers, the evolutions of social norms of privacy and control, and political values in the use of new technology for propaganda.

1.3. Research Questions

Our research will begin to unpack three important parts of computational propaganda from botnets: (a) the impact of bots on political discourse; (b) the differences between bot-generated and human posts; and (c) the transitional moment when bot-generated content is accepted and advanced by human users.

To date, what impact have innovations in auto-generated scripts on global social media services had on political discussions and current affairs? Who produces these scripts, or what are the conditions under which innovations in computer science and engineering get repurposed for "computational propaganda"? Is there a demonstrable impact of bots on news consumption? What is the evolutionary trajectory of this field of computer science, and what are the mechanisms for improving public literacy, generating careful policy oversight, and preventing the abuse of social networking technologies?

Despite the growing disbursement of this networked disinformation, little is known about the software's creation, dissemination, and capability. What does the presence and use of political bots mean for the young people who establish their political identity online and for those in authoritarian countries with restricted news media? How does governmental use of this new propaganda tool effect political organizing efforts and election outcomes? What does such information reveal about algorithmic culture, state interference in digital networks, and digital politics at large? Our proposed EAGER project engages with questions such as these via the construction and analysis of a globally comparative event dataset of political bot usage.

2. Research Plan

To answer these questions, we propose a three stage research design in which we build from grounded knowledge, qualitatively gathered, through an original comparative event dataset, to theoretical concepts that advance our understanding of how to track bot activity both socially and computationally. In part, the selection of field sites will be driven by the course of international crises over the next 24 months. In our team conversations we hypothesize that elections in particular are likely to involve the production of

politicized social media bots. Likely sites for such activity include the Syrian presidential elections in June 2014, the Turkish Presidential elections in August 2014, the Brazilian Legislative elections in October 2014, the UK general elections in May 2015, the Mexican Legislative elections in July 2015. Tracking bot production during these sensitive moments—and being responsive to other international crises as they develop—will help create the snowball sampling technique on bot producers.

2.1. International Fieldwork

The first and early stage of our collaborative project involves rigorous social science fieldwork methods to gather knowledge of how bot designers operate professionally, both in terms of working as an innovative network of engineers and as a professional network competing for clients in a market for computing services. PI Howard has a demonstrated record of working with hackers, hacktivists, spammers, and political campaign managers whose work violates most people's privacy norms and technology values. There are three target groups for our interviews:

(i) ***Makers of Political Bots***. These individuals and firms are internationally distributed, and some of the engineers behind political bots actually work at major advertising firms. The process of making contact and confirming willingness to participate in our interviews has already begun.
(ii) ***In-House Engineers***. All of the social media firms with high profile social media services employ computer engineers to detect bots. The process usually involves some algorithmic identification of problematic accounts that publish too quickly, but with the growing sophistication of bots human confirmation is often needed before accounts are sanctioned or deleted.
(iii) ***Industry Research Computer Scientists***. There are a few third party organizations dedicated to identifying bots. A few work for online services such as Status People and Truthy, but there are small research teams at Microsoft Research and UIUC that we have easy access to.

Contacts within the first two groups have already been made. We have already begun grooming network ties to the makers of political bots and the in-house engineers. Since one of the broad impacts of this project will be to improve the efficiency of bot detection and raise the shared understanding of how bot detection systems should evolve, and since we have our own professional ties to university-based researchers, contact with the third group will be made early in the performance of this project.

This component will take advantage of GSP Woolley's skills as an ethnographer and GSP Abokhodair's experience with qualitative information science research methods. Together they will interview bot makers who produce the automated scripts for political actors. Several are based in Russia, Bahrain, Hungary, and other countries within Eastern Europe and the Middle East. PI Howard has begun contact overtures and has a demonstrated record of being able to work ethnographically with the hacker community. GSP Woolley has contacted several bot makers in the US who are eager to connect the team to the wider international community and provide information themselves. GSP Woolley has also contacted bot makers in Eastern Europe and Russia. GSP Abokhodair has in depth knowledge of the Syrian bot situation and intends to contact Middle Eastern bot makers. Her fluency in Arabic will aid in this cause. Additional graduate assistants with proficiency in Mandarin, Russian, Ukrainian, and Spanish will aid in continued contact with bot makers in key countries that use these languages. Important nodes in the networks of political bot makers have already been located in Bahrain, Budapest, Moscow, San Francisco, Seattle, and Moscow. Important nodes in the networks of political bot detectors have been located in Seattle, London, and Urbana.

2.2. Comparative Event Dataset

The second involves the construction of an original event dataset more comprehensive than any previously collected. Event datasets have become particularly powerful tools for understanding trends in socio-computational systems, including global digital activism, compromised personal records, and government interference with digital switches (Edwards, Howard, and Joyce, 2012; Erikson and Howard, 2007).

To create this data set, a group of trained and supervised graduate student coders will review news stories created by both citizen and professional journalists which describe the impact of bots on political discourse. We already know of several dozen cases, and will use a purposive and snowball sampling technique

to identify cases. Research assistants will read each source and assign values for qualitative and quantitative variables defined in a crafted codebook. Our perspective will be global, and our objective in this research is to build a typology of the evolving use and impact of what we are calling "computational propaganda" in both democracies and authoritarian regimes around the world. This part of the project will involve both PIs and both GSPs, and will render important insight into the size of the consulting industry that produces political bots, and identify the key network actors we need to contact. Case coding will follow the high standards of these datasets and involve a small team of specially trained coders who participate in training, collect cases, research details, enter case information, participate in retraining, and get evaluated through inter-coder reliability scores.

Following the classical methodology of the study of unusual phenomena in technology diffusion and usability, we will begin our sampling of bot use by means of news reports about them (Earl et al., 2004). The media-based approach to data collection is especially valuable when the phenomenon at hand is particularly new. Our method for analyzing these texts will be a content analysis, a systematic means of textual analysis which endeavors to have all observers come to the same conclusions about the content of the text. This inter-coder agreement increases the reliability and also the authority of the attendant findings (Krippendorff, 2004).

Unlike many content analyses, in which unit of analysis and unit of observation are one in the same, in this study the two will be different. The unit of analysis is the bot campaign while the unit of observation is the news report about that campaign. This means that this stage of our analysis involves the indirect study of computational propaganda and gives us the added value of using third-party sources that can be evaluated for trustworthiness. Media bias is a legitimate concern that can be mitigated by relying on a variety of news outlets and on amateur as well as professional sources. This method alone will not completely nullify media bias, but combining the sampling strategy with knowledge garnered by international fieldwork with bot makers and bot detectors will help significantly. While traditional news sources and peer-reviewed journal articles will provide some entries, we expect that the interviewing process will drive our snowball sampling strategy to such a degree that we will come close to gathering the universe of known cases in which any political-motivated bot provided was designed and released on social media.

2.3. Tracking and Modeling Botnet Emergence

There are disparate projects dedicated to locating and detecting bots. The interviews with bot creators will help our team gain deep insight into three critical aspects of bot generation: a) What are the methods they use to influence online conversations? b) What hashtags/events do they target? c) How do they make bot posts similar to human-like posts? The interviews with the bot trackers – or the digital detectives as they like to be named—will help us better understand the methods used to track bots and learn how to control the domains that bot creators use to infect computers. After gaining insights from the two different groups of experts our plan is to start prototyping our own bots using the same techniques. The inner knowledge we as a team will gain from creating and disseminating bots will enable us to produce insightful recommendations on how to protect free speech and detect social bots.

In this third stage of our work, conducted after fieldwork and the construction of our event dataset, we will seek to elaborate and extend existing models of bot detection. This third piece of research will seek to account for the qualitative findings from the prior two components to model when computational propaganda will be used, and test the model by tracking and targeting Twitter and Facebook bots through targeted datasets that we assemble politically sensitive events. With their information science background CoPI McDonald and GSP Abokhodair will lead this stage, though this will also be an opportunity to teach PI Howard and GSP Woolley about the information science toolkit for such analysis. While much of the innovation here will be driven by our progress in the first and second stage of this EAGER project, it is likely that we will be taking advantage of context-relevant crowd sourced knowledge, existing scripts that CoPI McDonald has for Twitter analysis, and the additional computing skills of consultants who are expert in other social networking applications.

2.4. Pilot Activities and the Formation of a New Team

Our group conversations have evolved through sharing stories and solving problems on pilot activities that we had started separately only a few months ago. CoPI McDonald and GSP Abokhodair has been tracking

the use of bots by combatants in the Syrian Civil War. PI Howard and GSP Woolley have begun collecting cases for the event dataset of news reports of the political use of bots, and begun a cross national comparison on the use of bots in elections for a chapter contribution in an edited book (Woolley and Howard, forthcoming).

We have found our preliminary conversations over the last few months to be very productive, but our first real group conversations have been over the formation of this new research agenda and our new EAGER proposal. GSP Abokhodair has helped raise our understanding of how the automated scripts work, GSP Woolley has helped raise our understanding of the social context and impact of these scripts. So as a team, we now have a political ethnographer, a coder, and experience managing funded research and disseminating results. We seek the resources to formally combine our efforts in an extended, 2-year research plan. CoPI McDonald has been working with GSP Abokhodair on the problem of detecting bots as they go to work on high-volume hashtags.

This team pairs two senior faculty with two junior scholars, and two social scientists with two information scientists. Combined, our analytical toolkit includes social scientific and computational skills. While we have not collaborated before as a team, our prior dyadic associations and the healthy process of developing this research proposal together make us confident that we can see what we propose through to completion. This team will be working together for the first time, but PI Howard has previous experience managing teams of social and computer scientists on research into the political impact of innovation in science and engineering (RAPID IIS-1144286, IIS-0713074, ITR-0326101). Treating computational propaganda seriously—as a potentially transformative political problem—is a radical move for the social sciences. Studying innovation with event datasets and ethnography is rarely done in the computer sciences and engineering. So these different approaches and new team is likely to yield high impact insight.

3. Management Plan

We have a three pronged approach that takes advantage of the interdisciplinary skills in the team. We see the fieldwork as being an early-project activity. We would begin the event dataset after getting into the interviews because through those interviews we can begin to collect the case list for our event dataset. Our group conversations about tracking and models for detecting bots can begin as soon as we have some substantive field notes to drive our team dialogue forward. Table 2 details the pacing of work over the life of this EAGER project.

Table 2: Workflow, Tasks, and Roles from 2014-16

	Fall 2014	Winter 2015	Spring 2015	Summer 2015	Fall 2015	Winter 2016	Spring 2016	Summer 2016	Fall 2016
1. Prior to Award									
a) Apply for human subjects approval									
b) Plan and prioritize network contacts									
2. International Fieldwork									
a) Background research on the makers of political bots, in-house engineers, and the research computer scientists working in industry	■	■							
b) Develop interview script and participant observation plan	■	■							
c) Conduct fieldwork		■	■	■	■				
d) Exit fieldwork: observational memos, summary surveys				■					
3. Comparative Event Dataset									
a) Develop coding instrument, train coders, run pretest			■	■					
b) Code incidents for dataset				■	■	■			
c) Begin trend analysis						■	■	■	■
4. Tracking and Computational Theory									
a) Review fieldwork and event data for trends					■	■			
b) Develop socio-technical model for deploying a bot detector			■	■	■	■	■		
c) Deploy and test prototype bot detector							■	■	■
5. Research Dissemination									
a) Conference presentations				■	■	■	■	■	■
b) Article submissions							■	■	■
c) Release event dataset									■

The mentoring activities will occur at every stage of the project, and our engagement with scholarly conferences can begin mid-way through the project when we have enough interview and event data to be able to report results.

3.1. Response to Feedback on Summary Proposal

Feedback on our initial summary proposal indicated concern that the population of bot makers

would be difficult to reach. PI Howard has a demonstrated record involving deviant computer professionals and engineers in research. And as a team we have already located several potential research subjects (contact with them will be resumed with Human Subjects approval has been awarded). But in response to this feedback we have considered additional types of data collection to augment the data collected among the bot-makers. This project will be made even more feasible by adding other professionals working on the bots—the corporate engineers working for social media firms and the research staff of industry labs who are dedicated to improving bot detection.

3.2. Success of Prior NSF Support

Previous NSF support (IIS-0713074, $341,963, 2007-2010 and ITR-0326101, $1.23 million, 2003–2005) allowed the PI to investigate new ways of measuring the impact of the engineering standards setting process on technology diffusion in Central Asia, Tanzania, and the Middle East. PI Howard's "RAPID - Social Computing and Political Transition in Tunisia," (IIS-1144286, $45,625, 2011) allowed for additional fieldwork in Tunisia during that country's first real election and their simultaneous efforts at developing open telecommunications standards and less restrictive information policy. NSF support has enabled PI Howard to become one of the foremost experts on the political and policy process that can turn innovations in computer science and engineering into tools for social control (Hussain and Howard 2012; Hussain, Howard, and Agarwal 2011). NSF support has been acknowledged at in multiple articles and research monographs, including *Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam*, which is acknowledged to be the most prescient work on the role of ICTs in the Arab Spring and *Democracy's Fourth Wave? Digital Media and the Arab Spring* the first major manuscript on the role of social media in those popular uprisings (Howard 2010, 2013, 2014).

4. Justification for EAGER Support

It is very likely that foreign governments and political actors are already using bots to manipulate public opinion in the United States. This project will greatly expand our understanding of how this is done, and advance the conversation among researchers in the computing, engineering, and social scientists about the size of the problem and the possible solutions. But doing this well means taking advantage of our newly formed cross-disciplinary team.

Our project will build a prediction model of bot usage in upcoming international elections. In the next calendar year there are 17 highly contested international elections. Several of these are taking place in countries with authoritarian regimes and emerging democracies. Our project will work to predict political bot usage in these upcoming elections and determine what potential impact such use has on electoral outcomes. This project will build and make use of an original comparative "event dataset" that codes for bot sourcing, characteristics, and impact. Some of the most exciting and innovative tech research comes from ethnographers who study the makers and users of technology (Coleman, 2013; Beyer, 2014). Qualitative research of technology and technology makers and users provides crucial insight into human experience and cultural context, but also presents the possibility for the creation of innovative communication methods in the tech world and beyond. When integrated with computer science driven big data and quantitative research, methods like ethnography provide a keener, more user experience driven, approach to understanding and innovating a sector of society driven by data. Simply put, ethnographic research of technology provides a link from personal experience to machine prediction. We will study bot-making with bot makers, learning the ins and outs of the creation and deployment of the software from expert creators, with the intention of using garnered information to generate greater worldwide public understanding of computational propaganda. The data and understandings gathered from this empirical investigation will inform the third part of our study: targeted at combining evolving bot detection procedures with innovative crowd sourcing techniques to generate further holistic social and computer scientific comprehension.

We will interview bot-makers in key countries in order to generate insight into the inner-workings of political bots and botnets. This information will allow keen understandings of the nuances of hashtag usage, back-end social media site construction, bot mechanics,

Our research will demonstrate how bots impact the social systems in which they are deployed and how specific aspects of computational propaganda, data used for coercion, discrimination, and control, play

out globally. Such research will potentially generate new theoretical understandings and will certainly uniquely contribute to the standing theory of a variety of fields associated with the computer and social sciences.

5. Broad Impacts and Intellectual Merit

Through our new collaboration on this EAGER project we will be able to blur the boundaries between social and information science as it is traditionally practiced. We will combine qualitative fieldwork, comparative event data, and information science as research methods, and trace the evolution of new features of automated scripts for use in social network applications. We will map the movement of ideas and innovations across the ecosystems of bot makers, political actors, and bot detectors. And by working together, we can ground our observations of bot and human behavior in real-time and produce a rigorous event dataset that can serve the broad community of researchers working on the problem of bot traffic.

5.1. Broad Impacts

The event dataset produced by this research will have a broad impact on the network of industry and university researchers working on the problem of computational propaganda. Indeed, while our EAGER proposal will result in new computational theory about bot tracking based on our grounded study of bot producer networks, we are certain that the dataset will have a broad impact on the network of researchers working on detection. The cleaned dataset and codebook will be specifically shared with the teams of bot detectors who participate on our study.

Our team is after specific evidence of how learning, design and repurposing occurs among bot makers, not simply an archive of bot features. Second, timely research on the bot activity can best serve US foreign policy experts, computer scientists—and democracy—now. The State Department needs greater literacy on the impact of innovations in science and technology on politics, beyond what pundits provide. Social science research on human computer interaction is now extremely relevant for foreign policy.

5.2. Intellectual Merit

Bots and automated scripts do not simply burden digital networks. There is growing evidence that they can shape public opinion, but on their own social science and information scholars have not addressed this phenomena in a collaborative way. This work will help foreign policy makers better understand the relationship between technology diffusion and political processes. Will competitive elections and open political discourse, much of it mediated through mobile and digital technologies, produce more and more bots that work for social control instead of deliberative democracy? Ultimately, this project will have intellectual merit because it will produce the first standardized records of bot activity for future analysis. We cannot be certain about how bots will constrain or incapacitate social networks during important political moments over the next two years. But what is certain is that new norms of interactivity and expectations for information access are being encoded in these automated scripts, and public leaders in both democracies and authoritarian regimes are imagining new ways of shaping public opinion in ways most users do not fully understand.

The work is urgent because the number of bots seems to be growing, and their sophistication seems to be improving—especially so for the bots that are derived with a political agenda in mind. EAGER funding is the most appropriate mechanism for supporting the proposed work because new practices of social computing are being politically institutionalized now, and quick support will allow a new team of social and information scientists to focus on this unusual moment of political transition. This EAGER support will allow for rigorous social science in "real time." The intellectual merit of this proposal is in advancing our understanding of how contemporary political discourse works in socio-computational systems. Even if democratic practices do not prevail over the social media we study in particular countries, there is intellectual merit to explaining how modern publics conduct political conversations, in engagement with or ignorance of bots.