Proceedings of the Nineteenth International AAAI Conference on Web and Social Media (ICWSM 2025)

Do Bots Do It Better? Analyzing the Effectiveness of Automated Agents in State-Sponsored Information Operations

Cole Polychronis, Marina Kogan
University of Utah
cole.polychronis@utah.edu, kogan@cs.utah.edu

Abstract

State sponsored information operations, or SSIOs, are a growing problem across many of the information spaces we inhabit online. These instances of coordinated misinformation and propaganda have been perpetrated by over 80 state actors in the last decade, and have been used to exert influence on digital media consumption habits, discussions of contentious issues, and even national elections. Concern over the power that SSIOs wield is only growing as the proliferation of automated tools and services is making it easier than ever to launch large-scale manipulation campaigns. But what role do such automated agents play within the broader operations that they are deployed in? Are they even successful at making an impact in information spaces online? In this work, we address both of these questions through the use of a sequence-based clustering method and advanced linear modeling. Using these methods, we investigate the relationship between agent automation, role, and network characteristics and how much success those agent's achieve over the course of their lifetimes. We find that automated agents perform worse across every success metric compared to human agents, and that they play a smaller, supporting role to the primarily human SSIO workforce. What's more, we find that the extent to which agent's engage in amplifying- or producing-centric roles is by far the biggest determinant of how successful they will be, highlighting the importance of social-roles in the analysis of automated agents.

Introduction

During the 2016 U.S. presidential election, an office of over 1,000 operatives posed as U.S. citizens and produced divisive content across multiple social media platforms, with the end goal of influencing the results of the election (Lister, Sciutto, and Ilyushina 2017). This was the work of the Russian-employed Internet Research Agency (IRA), and by the end of this campaign, over 126 million Americans were exposed to IRA-produced content on Facebook alone (Ingram 2017). Such campaigns, which we refer to as state-sponsored information operations, or SSIOs, have only become more common in the last decade. Recent reports indicate that just in 2020, over 80 state actors engaged in *at least* one attempt to manipulate public opinion online (Bradshaw, Bailey, and Howard 2021). Concern over the influence that SSIOs wield on public information spaces is only growing as large language models continue to improve and proliferate. This proliferation is lowering the barrier to entry for state actors, private firms (Bradshaw, Bailey, and Howard 2021), and even individuals (Knight 2023) to perpetrate information operations. But are such automated agents actually that effective at manipulating public information spaces? Recent work from Cresci et al. has shown that a common motivation for studying such automated agents, which they refer to as social bots, is an assumption that automated agents are largely responsible for manipulating information and exacerbating social conflicts, even though existing literature paints a far murkier picture (Cresci et al. 2023). They find that most literature focuses solely on detecting and characterizing the behavior of of such automated agents, leaving the question of what role these accounts play, and critically, how effective they are, largely unexplored.

In this work, we quantitatively address both of these questions in the context of automated agents used by SSIOs operating on 𝕏 (formerly Twitter). Specifically, we apply recent developments in social role identification (Polychronis and Kogan 2023) to measure the utilization of automated agents in SSIOs *and* what roles they serve. We then use Generalized Linear Mixed Modeling to measure what effect this utilization (both at the operation- and individual-level) has on the effectiveness of human and automated SSIO agents. Through our analysis, we find that when the role of an agent is not considered, automated agents appear to present a tradeoff of providing better engagement, at the cost decreased "staying power" (less followers and a shorter longevity). However, once we account for role, we find that automated agents perform worse on *every* success metric when compared to human agents. Once we account for role, we also find that an agent's role, and specifically whether they amplify or are being amplified by other agents, is a far bigger factor in agent success than whether or not that agent is automated. We conclude this paper with a discussion on the importance of social roles in automated agent research, a reflection on the limitations of our work and the directions these provide for future work in this domain.

Related Work

Below, we introduce definitions for State Sponsored Information Operations and social botnets, and then discuss how social roles manifest in the SSIO context.

State Sponsored Information Operations State Sponsored Information Operations (SSIOs) can broadly be understood as concerted efforts to manipulate social media in order to "manufacture consensus, automate suppression, and undermine trust in the liberal international order" (Bradshaw and Howard 2019). SSIOs typically evolve in three stages, first creating and utilizing accounts to mimic authentic users and gain legitimacy, before using coordination to generate, amplify, and spread their own content. If SSIOs are successful, the third step, then, is that genuine users are convinced of the value of this content and spread it themselves (Luceri et al. 2024). Originally, SSIOs were primarily conducted by governments, state actors, or political figures (Bradshaw and Howard 2019), but increasingly the work of conducting SSIOs is being outsourced to private firms, with at least 48 such firms being hired in 2020 (Bradshaw, Bailey, and Howard 2021). While the goals of SSIOs range from inciting social tension to quashing civil unrest (Cordey 2019), they all primarily operate by weaponizing various forms of information, whether that be propaganda, or the intentional (disinformation) or unintentional (misinformation) spread of false information (Theohary 2020). Over time, the strategies and techniques employed by SSIOs have become more accessible (Knight 2023) and more readily adopted by non-state actors (Beers et al. 2023). In this work, we consider one of the overarching strategies used by SSIOs, role-taking (Polychronis and Kogan 2023; Cima et al. 2024), and analyze how this strategy contributes to operation success.

Social Bots Defining a social bot is surprisingly difficult, as social bots are studied in a variety of contexts and by a multitude of communities. The works that have aimed to bridge these different context and communities tend to agree that "social bots" are social media accounts that exhibit either full or partial automation, which are typically deployed in groups (called social botnets), that strive to mimic human behavior to avoid detection and manipulate public opinion (Cresci 2020; Gorwa and Guilbeault 2020; Grimme et al. 2017). While early social bot research mainly considered things like fake followers and simple spam-bots, the theories and detection techniques developed in the social bot domain were also applied with great effect to early SSIOs (Hegelich and Janetzko 2016; Abokhodair, Yoo, and McDonald 2015). However, the many different contexts in which the social bot term is applied, as well as the way that social bots are constantly evolving in order to avoid detection (Cresci 2020), results in a key tension in social bot research: not all social bots are the same. More specifically, social bots can behave very differently depending on what they are built/deployed to do and as such, this context must be considered deliberately (Cresci et al. 2023). In this work, we attend to this consideration by specifically scoping our analysis to the social bots participating in SSIOs, which we refer to in the remainder of this paper as automated agents.

Social Roles in SSIOs Coordination is at the heart of the work that SSIOs undertake. Agents can be coordinated around what topics they should post about, which people they should target/amplify, what platforms to use, and in at least one case, what profanity/derogatory language they can use without breaking community guidelines (Seddon 2014; Broderick 2019). Recent work from Polychronis and Kogan has shown that one of the other significant ways that this coordination manifests in SSIOs is in the social roles that different agents occupy (Polychronis and Kogan 2023). These social roles allow individuals to contribute to complex tasks and achieve goals, even without being fully aware of those goals (Weick 1993; Yang et al. 2016). Polychronis and Kogan used social roles to investigate how principles and detection techniques from social bot research were being applied in this SSIO context. In this work, we more deeply explore the nature of these social roles, specifically to investigate differences in the role that human and automated agents serve in SSIOs and to what effect

Methods

To answer our questions concerning the role that automated agents play in SSIOs, and how effective they are, we turn to the 𝕏 Information Operations archive. We then use embedded Digital DNA, a sequence based clustering technique, to extract characteristics related to agent automation and role. Finally, we combine these and several other agent characteristics, as well as several operation-level traits, and perform regression to explore the relationship these characteristics have with the success that an agent achieves. We describe this dataset and these modeling techniques below.

Data

To explore the role taking and efficacy of SSIO agents on 𝕏, we analyze data from their Information Operations archive[1]. At time of analysis, this archive contained the complete activity of agents tied to 34 different operations. However, after classifying agents as being automated or human (described in the following section), we find that nine of these operations make no use of automation. As our analysis focuses on comparing human and automated agents within and between SSIOs, we remove these nine operations from consideration, leaving us with a final dataset of 25 operations. In total, this dataset contains over 207 million posts, authored by 78,151 agents. One of the critical data-cleaning/processing steps which 𝕏 takes before disclosing SSIO agent data is that for engagement and following metrics (likes, reposts, number of followers, etc.), they *remove* any engagement that comes from suspended accounts (which would include any engagement that comes from within the operation itself). For example, a post from an agent in this dataset with 100 likes, of which 90 came from other SSIO agents, would be recorded as only having 10 likes. The biggest limitation of this dataset is that any agent with less than 5000 followers, or any mentioned user with less than 5000 followers, is anonymized. Between this and the fact that all SSIO agents present in the archive have been suspended from the platform, additional information, such as social network characteristics or engagement metrics including the inflation due to other SSIO

[1]The 𝕏 Information Operations archive can be found at: transparency.x.com/en/reports/moderation-research.html

agents, can't be accessed with the 𝕏 API. Despite these limitations, this is one of the largest, and most under-explored (Cresci 2020) SSIO datasets publicly available, making it a perfect fit for our analysis.

Identifying SSIO Agent Automation and Role

In order to assign roles to SSIO agents, we use the Digital-DNA methodology developed by Cresci et al. (Cresci et al. 2016, 2017) and extended by Polychronis and Kogan (Polychronis and Kogan 2023). This method works by encoding the individual activities of users (in this case, posts on 𝕏) using one or more mappings, referred to as alphabets. Through this process, we convert user activity timelines into strings of characters, which we can then embed numerically and cluster. These clusters can then be summarized in order to reveal the social roles that agents within a cluster serve. While Polychronis and Kogan used this method to understand agents in the SSIO context, they identify that a limitation of their approach is that the alphabets that they used to encode activity were largely informed by the social bot literature, and that further analysis in the SSIO context would benefit from a more SSIO-specific set of alphabets. To this end, we propose a modified set of three alphabets, shown in Figure 1 and described in greater detail below, that draws more directly from the SSIO literature.

$$\alpha_1 = \begin{cases} A \leftarrow & \text{standard tools used} \\ B \leftarrow & \text{automated tools used} \\ C \leftarrow & \text{other tool-types used} \end{cases}$$

$$\alpha_2 = \begin{cases} D \leftarrow & \text{content production} \\ E \leftarrow & \text{content amplification} \end{cases}$$

$$\alpha_3 = \begin{cases} F \leftarrow & \text{short time between activities} \\ G \leftarrow & \text{long time between activities} \end{cases}$$

Figure 1: The SSIO-informed alphabets we use to encode agent behavior. Every tweet is described using three characters describing the tool used to author the post (α_1), the type of post (α_2), and the time since last activity (α_3). These character strings are then combined for each agent using embedded digital-DNA to assign each agent a role.

Account Automation To encode account automation (α_1), we use a similar process as Polychronis and Kogan (Polychronis and Kogan 2023). We find that their methodology for constructing this alphabet, while informed by the social bot literature, mirrors the current understanding of what suites of tools are employed in SSIOs (Bradshaw and Howard 2019). The first author manually encoded the tool-type used to author a post as either human-oriented, automation-oriented, or custom-built tools. Human-oriented tools are those that exist in the native 𝕏 ecosystem, such as the mobile and web clients, while automation-oriented tools are those that explicitly advertise automation or scheduled-posting as one of their features. Custom-built tools are those that either do not advertise such features or those that are not advertised online in any capacity. We exclude agents from our analysis that primarily use custom tools, as it is too difficult to determine at scale whether these tools are simply niche services (which would suggest that agents are humans) or if they are tools built by SSIO organizers to automate work (Cryst et al. 2021; DiResta, Grossman, and Miller 2019). We find that the vast majority of operations have no agents who primarily use such custom-built tools (as shown in Figure 2) and thus feel comfortable excluding them from our analysis.

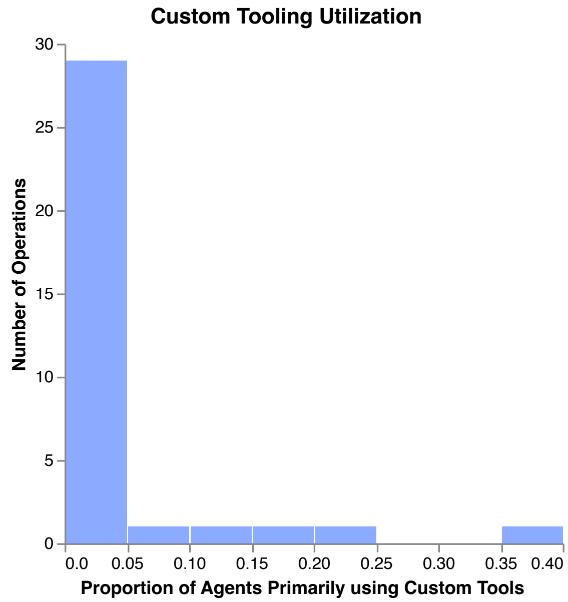

Figure 2: The proportion of agents that primarily use custom tools. Only five operations make greater than 5% utilization of custom-tooled agents.

Activity Type To encode activity type (α_2), we operationalize the well-established producer-amplifier relationship that has been identified in an array of SSIOs, including ones active in Syria (Abokhodair, Yoo, and McDonald 2015), Germany (Neudert, Kollanyi, and Howard 2017), and the United States (Arif, Stewart, and Starbird 2018). In order to do this, we encode agent activity into a binary, and classify any activity which produces new content on 𝕏 (i.e. posting and replying) as "production activities" and any that amplify/propagate existing content, in this case reposting, as "amplification activities".

Activity Frequency To encode our final alphabet (α_3), we encode the time elapsed since last activity as being either short if it was on the order to second or minutes after, or long for activity occurring hours after or longer. This binary encoding is quite simplisitic, but captures when agents engage in bursty activity with zero or near-zero downtime between posts. This burstiness has been identified as a key feature of social media manipulation by both the social bot literature (Mazza et al. 2019) and the SSIO literature (Abokhodair, Yoo, and McDonald 2015).

Measuring Agent Effectiveness

In order to investigate the impact that agent automation/role has on the success they achieve, we perform regression to measure the relationship between agent characteristics (both at the individual- and operation-level) and several common success criteria (one regression model per success metric). To emphasize, we are measuring success at the *individual agent* level, with an assumption that an operation's success can be seen as the aggregate of the success that its agents achieve. To perform this regression, we utilize a class of models called Generalized Linear Mixed Models (GLMMs). GLMMs are simple extensions of standard linear/hierarchical models that provide two key benefits. First, they accommodate response data that is not continuous, such as binary or count data (Skrondal and Rabe-Hesketh 2003). Naturally, this characteristic is important because the success criteria that serve as our response variables are counts, rather than continuous values. Second, GLMMs allow for the use of random effects to account for correlated observations (i.e. when some samples come from the same group) (Bolker 2015). This feature is useful to account for baseline differences in performance between operations, which may obscure the difference between automated and human agents. In particular, this quality is an important means of controlling for the fact that there are varying levels of use of automation between different SSIOs in our dataset.

In this analysis, we use an even more specialized GLMM, the Zero-Inflated Negative Binomial (ZINB) model[2]. In addition to the benefits listed above, ZINBs are uniquely suited for this analysis because they work well with response variables that are not normally distributed – particularly if they are binomial or zero-inflated/dense in nature. It is well established that social media metrics (including our success metrics) are highly skewed towards zero, making this feature of ZINBs critical (Kwak et al. 2010). Below, we discuss the success metrics that make up our response variables and the agent- and operation-level metrics that make up our explanatory variables.

Response Variables We utilize the following three metrics for measuring the success of SSIO agents: engagement, follower count, and longevity. **Engagement** combines the total number of likes, reposts, and replies that an agent receives over their entire lifespan and is the most straightforward success metric of the three. In addition to engagement capturing the popularity of agent-produced content, engagement also serves to algorithmically boost content so that it is visible to more people. Getting 'real' users to propagate SSIO-produced content, even if unwittingly, is the ultimate goal of most SSIOs (Bittman 1985). **Follower count** works in a similar way, and serves as a measure of how many real users are opting-in to receiving more content produced by SSIO-agents, which in turn makes them more likely to propagate it (Kwak et al. 2010). Finally, **longevity** is a natural success metric because greater longevity means more time to produce or amplify content, and longer to influence and manipulate information spaces (Bradshaw and Howard 2019). One caveat with longevity is that rather than measuring it as time elapsed from account creation to date of last tweet, we measure starting from the 5% quantile of posting (i.e 5% of the way into when an account starts posting). This is necessary as many SSIOs make use of hijacked accounts, or genuine accounts that were purchased, stolen, etc. by operation agencies to skirt suspicious account creation checks done by $\mathbb{X}$ (Bradshaw and Howard 2019). While this is by no means guaranteed to remove all of (or only) the activity produced by non-agents, it does serve as a way to normalize account activity to that mostly produced by agents.

It is important to note that these three metrics are not the only suitable measures of success. One could imagine that measuring change in opinion of real users or change in hashtag utilization to be equally valid measures of success. However, not every SSIO has the same goal; while constructive operations (which aim to establish coherent narratives) might very much care about changing people's opinions, disruptive operations (which aim to polarize crowds and foment distrust) often don't care if people's opinions change (Cordey 2019). We choose the metrics of engagement, follower count, and longevity as these will be relevant regardless of what an SSIO's goals are.

Explanatory Variables The explanatory variables that we build our regression on are shown in Table 1. We consider nine agent-level variables and four operation-level variables, as well as a random effect to account for operation membership, for a total of fourteen explanatory variables. **Agent Type** (drawn directly from α_1) is the most central variable in our analysis, allowing us to investigate the efficacy of automated agents in SSIOs. **Role Activity** and **Role Frequency** are drawn directly from alphabets α_2 and α_3, allowing us to investigate the impact an agent's role has on success. **Agent Focus** encodes the proportion p of an agent's total activity that is devoted to its two most frequent activities. In other words, this metric measures whether an agent is completely devoted to their primary role, or if they engage in a wide range of activities outside their primary role. We assign focus based on the following heuristic:

Clear Focus:	$p \geq 0.75$
Semi-Clear Focus:	$0.75 > p \geq 0.50$
Diffuse Focus:	$0.50 > p \geq 0.25$
No Focus:	$p < 0.25$

The next four variables in Table 1, relating to the **Inflow** and **Outflow** of agents, encode the amount of engagement that an agent gives and receives from human and automated agents, and allows us to more directly explore the amplification dynamic that was seen in early SSIOs (Abokhodair, Yoo, and McDonald 2015; Neudert, Kollanyi, and Howard 2017; Arif, Stewart, and Starbird 2018). The next variable, **Outflow to Real Users**, is an important complement to the previous variables, as no work has distinguished between the effect of amplifying content that originated from within an SSIO versus that which originated from outside of an SSIO.

[2]We tried several other GLMMs, including Poisson and Negative Binomial models. These did not converge due to the highly dispersed nature of our data and were removed from consideration per the recommendations of Bolker (Bolker 2015).

Individual-level	Agent Type	whether an agent primarily uses automated or standard tools
	Role Activity	whether an agent primarily engages in production or amplification
	Role Frequency	whether an agent primarily acts frequently or infrequently
	Role Focus	degree of focus that an agent has on it's main activity
	Outflow to Humans	amount of engagement an agent gives to human agents (log)
	Outflow to Bots	amount of engagement an agent gives to automated agents (log)
	Inflow from Humans	amount of engagement an agent receives from human agents (log)
	Inflow from Bots	amount of engagement an agent receives from automated agents (log)
	Outflow to Real Users	amount of engagement an agent gives to non-agents
Operation-level	Operation Size	number of agents involved in the operation (log)
	Number of Roles	number of unique roles undertaken by agents within an operation
	Automation Usage	proportion of agents in an operation that are automated
	Network Density	density of the engagement network between agents within an operation
Random Effect	Operation Membership	random effect to account for which operation an agent belongs to

Table 1: The individual-level, group-level, and random effect explanatory variables that we relate to agent success using ZINBs. In all, we use fourteen variables for each of our three regression models.

Operation Size, **Number of Roles**, **Automation Usage** and **Network Density** allow us to consider the amount of resources or logistical planning that are invested into an operation, which is an important dimension to consider given that SSIOs exist at the scale of a handful of accounts operating independently and at the scale of hundreds of thousands of accounts all amplifying one another (Bradshaw and Howard 2019). Finally, we include a random effect to control for which operation an agent belongs to in order to account for the fact that some operations are simply more successful than others because of the topics or communities they engage with or the high-level goals that they have (Cordey 2019). The operation-level variables, in conjunction with the random effect, also serve a double purpose of allowing our analysis to control for the fact that the SSIOs in our dataset are fairly heterogeneous, embodying a wide range of organizational structures, utilization of agents, and reliance on automation.

Findings

In the sections below, we explore the scope and uniqueness of roles taken on by the agents of SSIOs in our data set. We then present how an agent's role and related characteristics impact the success it achieves over the course of its lifetime, and how this differs between humans and automated agents. Finally, as social roles are not commonly considered in SSIO or automated agent research, we explore what happens when agent success is measured without accounting for the role that the agent serves.

Minimal Automation Usage in SSIOs

After characterizing SSIO agents using the embedded Digital-DNA approach described above, we find that automation is used rather sparingly in most operations, as shown in Figure 3. In fact, we find that over a quarter of the operations in our dataset make no use of automation whatsoever, and further that there are only six operations where automated agents make up a majority of the operation's workforce. Given that much of the motivation for studying automated agents in the SSIO context is an assumption that automated agents play a critical role in misinformation propagation (Cresci et al. 2023), their relatively light use in most operations is surprising. However, this result says nothing of the role or effectiveness of these automated agents, however lightly used they may be. We next turn to looking at the high-level roles that humans and automated agents play.

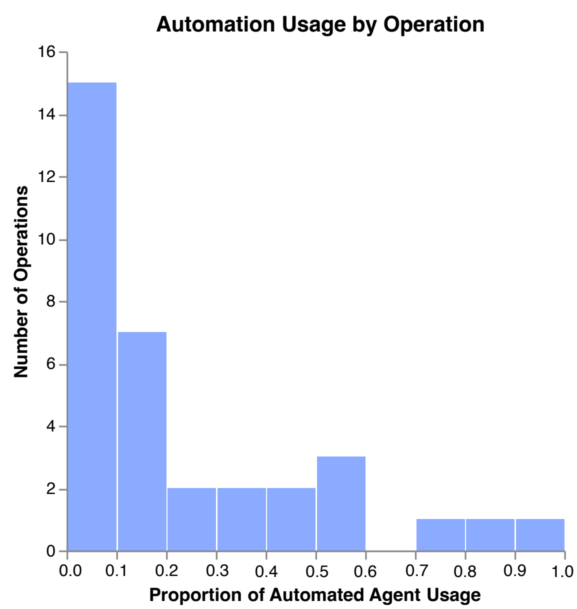

Figure 3: The proportion of automated agents used by SSIOs in our dataset. For a majority of these operations, automation makes up less than 20% of their workforce, with nine operations making *no* use of automated agents.

Automation Does Not Always Serve a Unique Role

In Figure 4, we show the proportion of automated agent roles that are already being served by human agents for each operation. We find that while eight operations fit the common misconception that automated agents are responsible for some critical, unique role (Cresci et al. 2023), most operations do not. In fact, at the opposite extreme, we see that there are five operations in which the roles served by automated agents are completely redundant to the ones that humans serve, and a further seven operations where there is at least a 50% overlap between the roles of human and automated agents. Taken together with our findings on the minimal usage of automation in SSIOs, this indicates that the role that automated agents serve in SSIOs is fairly limited (as they are deployed sparingly), and oftentimes, they are not even doing something that human agents weren't already doing. However, this still leaves the question of how much these automated roles (whether unique or not) contribute to operation success, which we explore next.

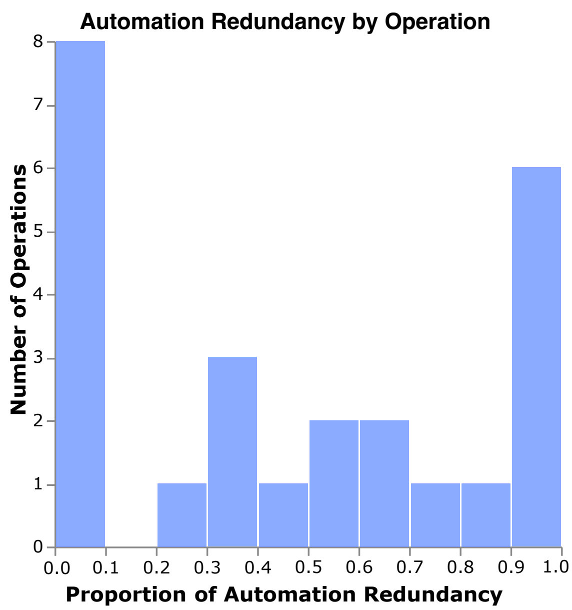

Figure 4: The proportion of automated agent work that is already performed by human agents. While we see automated agents serve completely unique roles in some operations, we see almost as many operations where the roles of automated agents completely overlap with those of humans.

Human Agents Outperform Automated Ones

Below, we present the results of our ZINB regressions on engagement, follower count, and longevity. All statistically significant ($p < 0.05$) results are shown in Figure 5 and full regression information is shown in Tables 2 and 3 in the Appendix. Overall, we find that automation is an important factor in agent success, but far more important is the trade-off in giving and receiving engagement from other agents – agents are substantially more effective when they receive engagement from others, but this comes at the cost of decreased effectiveness from the agents doing the engagement.

Agent Automation Starting with the most central finding of our analysis, and given trends in the prior literature, perhaps the most surprising, we find that automated agents actually perform substantially worse than human agents across all three of our success metrics (Figure 5A). In particular, automated agents struggle to garner similar followings as humans, generally having follower networks that are a third smaller than the networks humans attain.

Agent Role We find that agent roles are a far more important factor in agent effectiveness when compared to whether or not they are automated. In particular, we find that whether an agent focuses on amplification or production is a significant determinant of individual success – amplifying agents receive a third of the engagement and following and 40% shorter longevity then production-focused agents (Figure 5E). We also see that agents that are frequently active are significantly less successful than agents that spread their activity out (Figure 5F). These findings map quite cleanly to what has been found in the social bot literature - amplifier and/or spammer bots tend to make little impact and are quickly detected and acted against (Cresci 2020).

The impact that focus on a role has on agent success is more complicated. We see that unfocused agents (Figure 5D) perform significantly worse than clearly focused ones. In contrast, semi-clear or diffuse focus (Figure 5B-C) leads to improvements in success, with diffusely focused agents enjoying incredible gains in engagement. In large part, these findings support recent studies that have shown that adaptable agents, that employ a variety of behaviors, tends to be of better service to SSIOs (Bradshaw and Howard 2019). However, our results also present an important caveat – that agents that employ too many behaviors – and as a result are unfocused, perform worse than those with greater focus on a specific activity.

Agent Amplification Looking at our explanatory variables related to agent amplification (Figure 5G-J), we see the trade-off inherent to amplification that was touched on in the previous section come into sharp focus. From Figure 5G, we see that just a 1% increase (as these are log-transformed) in the amount of engagement an agent has with human agents results in a nearly 15% reduction in engagement and following, and a nearly 5% reduction in agent longevity. Interestingly, we see that these negative effects are less noticeable when agents amplify automated agents, with no significant impact on agent longevity. In contrast, we see that just a 1% increase in amount of engagement an agent *receives* from human agents yields a 33% increase in following and more than doubles the expected engagement, with a decent increase in longevity as well. We see this benefit of amplification hold for amplification coming from automated agents as well (Figure 5J), but the effects are less substantial, especially for engagement. Taken together, these results show an intrinsic tradeoff of agent amplification, where agents that are amplified experience significant boosts in success, while the agents that do the amplifying experience substantial de-

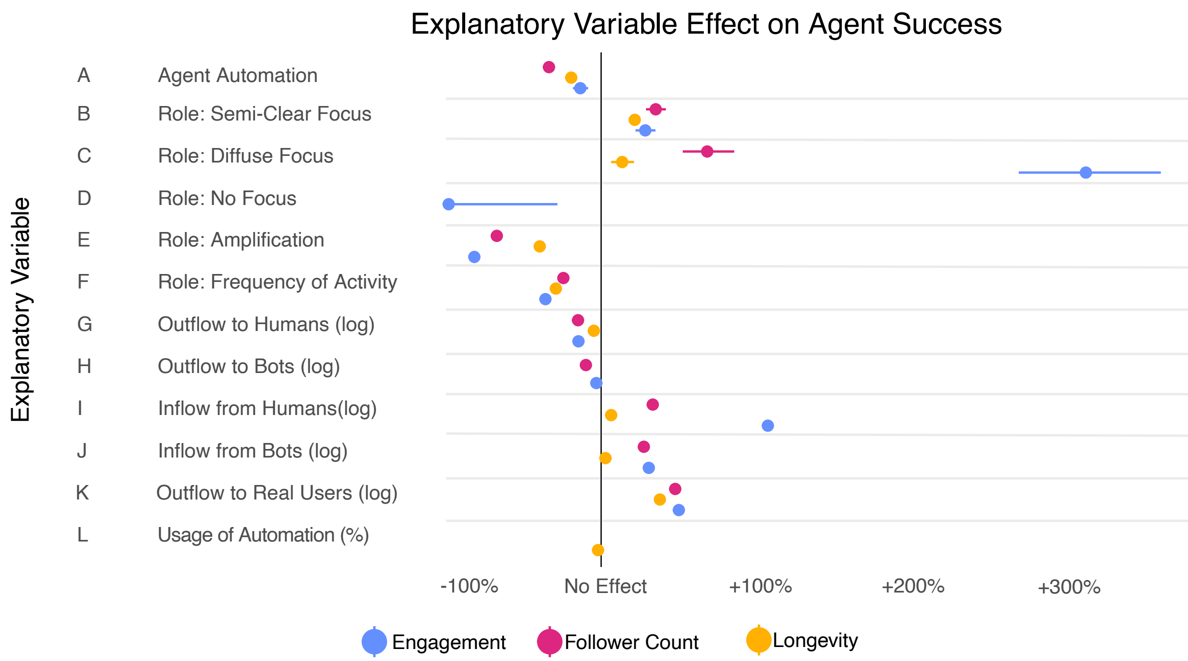

Figure 5: The impact of our statistically significant explanatory variables on engagement, follower count, and longevity. We show 95% confidence intervals of the predicted effect size (for many estimates, the bars showing these intervals do not appear because the interval is so narrow). Effect sizes are calculated as $e^{\theta} - 1$, where θ is the coefficient predicted by the GLMM regression. While automation and agent role have clear impacts on agent success, by far the most powerful explanatory variables are those that relate to the amount of amplification that an agent receives and gives.

creases in their own success. An important caveat to this trade-off is that it does not hold for engagement given to "real users" (Figure 5K). Here, we see that the more engagement an agent gives to real users, the greater success they experience. While individual case studies have shown the tradeoff in success between giving and receiving engagement (Abokhodair, Yoo, and McDonald 2015), our findings are the first to show this holds longitudinally across many SSIOs, and that it does not apply to amplification of users existing outside of SSIOs.

Operation-level Characteristics Overall, we see that operation-level characteristics play no significant role in agent success, with one exception. Further supporting our high-level finding that automated agents are far less successful than human ones, we see that an increase in the usage of automation (i.e. the proportion of agents in an operation that are automated) leads to a decline in agent success. Specifically, a 1% increase in the usage of automated agents leads to a 2% decrease in the overall longevity of SSIO agents (Figure 5L).

Roles are Critical for Understanding Agent Success

In order to consider the current state of SSIO and social bot research, which largely does not take role into account (Polychronis and Kogan 2023), we perform the same regression analysis as before, but without the inclusion of our role-based explanatory variables, with the exception of agent automation, as this is central to our research question. The full results of that analysis can be found in Tables 2 and 3 in the appendix, but we present notable findings in Figure 6. Overall, we find that the effect and significance of explanatory variables mirror the results of the original regression presented in Figure 5, with two important exceptions. First and foremost, we find that when we don't account for roles, automated agents are no longer strictly worse than human ones and instead present a trade-off of better engagement at the cost of cultivating a smaller following and living for a shorter duration. Second, we find that without accounting for role, the size of an operation becomes a negative determinant of success: the bigger an operation, the less successful its individual agents are predicted to be across all three metrics. In short, when we don't account for role, this collapses the effects that role plays in agent success, and makes automated agents appear more effective than they actually are. As roles are not commonly analyzed in the current slate of SSIO research (Polychronis and Kogan 2023), this effect-collapse offers one explanation for why determining automated agent effectiveness has been so difficult to triangulate (Cresci et al. 2023).

Discussion

Below, we present a summary of our findings, and the implications they hold for automated agent research in the SSIO context, before concluding with a discussion of the limitations of our work and the directions they provide for future work.

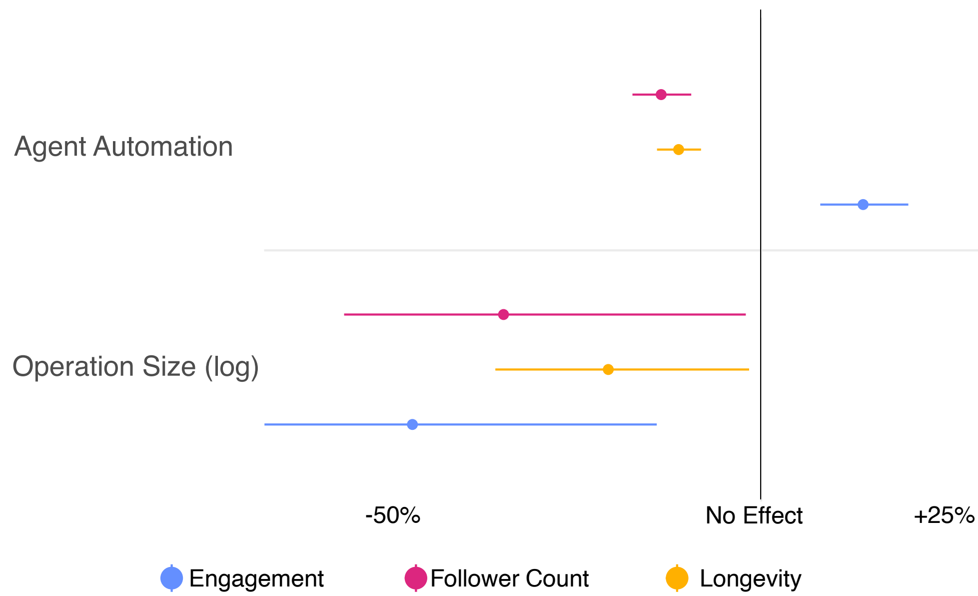

Figure 6: The impact of automation and operation size on engagement, follower count, and longevity when we do not account for agent role. While the relationship between many of our explanatory and response variables are similar between models that do and do not account for role, we see significant deviations with automation and operation size. Automation now presents a tradeoff in comparison to humans, rather than a strict downgrade, and operation size correlates with a decreased effectiveness of agents across all three success metrics.

Summary of Findings In this paper, we explore a common motivation for social bot research: that they are responsible for misinformation and social discord in online information spaces (Cresci et al. 2023). Looking at SSIO activity on 𝕏, we find that automated agents oftentimes play a smaller, complementary role to the work that human agents do. If we don't take roles into account, automated agents appear to present a tradeoff in effectiveness: yielding increased engagement at the cost of reduced staying power. However, when we analyze agent effectiveness with social roles in mind, we find that automated agents perform worse than humans across every metric. Moreover, we find that an agent's role is a more significant determinant of success than automation and in particular, the degree to which agents give or receive engagement factors massively into their success.

Implications In a provocative reflection on the many contrasting findings in misinformation research, Derek Ruths commented that this body of work "has come to resemble the very thing it studies" (Ruths 2019). Cresci et al. implore that if this trend is to be reversed, then more nuanced work that engages with the complexity of agents engaging in misinformation must be undertaken to disentangle and explain the contrasting findings that are found in this body of research (Cresci et al. 2023). Our analysis points to one possible source of the contrasting findings in misinformation research: a relative lack of consideration being paid to the social roles of those participating in misinformation. Our findings show that an agent's role has a considerable impact on their success, suggesting that if roles are not taken into account, significant uncertainty is introduced into what kinds of agents actually achieve what levels of success. Given agent effects and effectiveness are currently underexplored (Cresci et al. 2023), as more researchers begin to grapple with this topic, our findings indicate that an agent's role is just as important of a factor to consider as whether or not an agent is automated.

Our work also shows the utility of the embedded Digital-DNA method beyond the high-level analysis it was designed for. Embedded Digital-DNA has, up to this point, been used primarily to analyze SSIO agents at the group-level, being used to describe high-level SSIO organization and for aiding in the development of SSIO detection techniques (Polychronis and Kogan 2023). Through our analysis, we show that in addition to this high-level utility, the social roles returned by embedded Digital-DNA have significant utility at the individual level, serving as potent descriptors, particularly for the purposes of modeling.

Limitations and Future Research While our work represents significant progress towards understanding what automated agents actually contribute to SSIO success, there are several limitations to our approach that present opportunities for future research. One of the important factors of SSIO construction that we do not engage with in our analysis is the constraint of cost. Recent research on social bot misinformation and harassment has found that buying the services of a social bot can be one or even two orders of magnitude cheaper than paying the salary for a single human operative (Berman 2022). From this perspective, automated agents might present a real tradeoff, doing everything worse than humans, but at a steep discount. One of the key difficulties of accounting for cost in historical SSIO data is that SSIOs were most commonly funded out of the budgets of intelligence/counterintelligence agencies or entities of the state sponsoring the operation (Bradshaw and Howard 2019), meaning that most of this data is closely guarded even to this day. However, SSIOs are increasingly being run by third party, 'grey public relations (PR) firms' and not actual states or state agencies. Some of these PR firms even advertise their services and rates online (Gallagher 2019). Future work could triangulate automated agent cost estimates from these firms' listings, and introduce this as a constraint in order to investigate optimal configurations of SSIO agents.

An intrinsic limitation of our approach is that the alphabets that we choose to encode agent behavior with are largely drawn from previous work (Cresci et al. 2016) and are by no means exhaustive. In particular, the alphabets we utilize largely highlight *behavior* and as such, a valuable path for future exploration would be development of more content-focused alphabets to investigate the intersection of an agent's content and role, as well as whether agent content production is largely top-down (Seddon 2014; Broderick 2019) or more prone to individual improvisation. The content-based measures that Starbird et al. (Starbird et al. 2018) and DiResta et al. (DiResta et al. 2019) utilize in their studies of multi-platform information operations would be valuable both for studying SSIO content and for developing platform-agnostic alphabets.

Additionally, a current limitation of our approach is that agent automation is encoded as a binary variable – either an agent primarily uses automated tools or it does not. However, recent work from social bot research has identified that such binary classifications are often quite limited, and fail to account for the complexity and array of activities that these agents now undertake (Cresci 2020). This problem is only being magnified with the rapid development and adoption of large language models (LLMs), with LLM-powered accounts having been shown to be more effective than traditional social bots, while still being cheaper than human agents (Knight 2023). A powerful direction for future work, then, would be to expand automation classification out to more categories, or even to a spectrum (Davis et al. 2016), in order to gain even more nuanced insight into how such factors impact agent success.

Finally, we turned to the 𝕏 Information Operations archive to perform our analysis as it is one of the largest SSIO datasets currently available with significant potential to empower the "next wave of research" in this domain (Cresci 2020). A clear limitation of using this dataset is that it is unclear if our findings generalize to SSIOs run on other platforms. In particular, platforms like Facebook, TikTok, and WhatsApp have seen substantial and sustained growth in the number of SSIOs using their service (Bradshaw and Howard 2019) and work to apply a social-role informed analysis of agents effectiveness on these platforms would be particularly valuable in the SSIO research space.

References

Abokhodair, N.; Yoo, D.; and McDonald, D. W. 2015. Dissecting a Social Botnet: Growth, Content and Influence in Twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, 839–851.

Arif, A.; Stewart, L. G.; and Starbird, K. 2018. Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW): 1–27.

Beers, A.; Schafer, J. S.; Kennedy, I.; Wack, M.; Spiro, E. S.; and Starbird, K. 2023. Followback Clusters, Satellite Audiences, and Bridge Nodes: Coengagement Networks for the 2020 US Election. *Proceedings of the International AAAI Conference on Web and Social Media*, 17: 59–71.

Berman, V. 2022. Trolling as a Service in 2022.

Bittman, L. 1985. *The KGB and Soviet disinformation: an insider's view*. Pergamon-Brassey's. ISBN 978-0-08-031572-0.

Bolker, B. M. 2015. Linear and generalized linear mixed models. In Fox, G. A.; Negrete-Yankelevich, S.; and Sosa, V. J., eds., *Ecological Statistics*, 309–333. Oxford University PressOxford.

Bradshaw, S.; Bailey, H.; and Howard, P. N. 2021. Industrialized disinformation: 2020 global inventory of organized social media manipulation. *Oxford Programme on Democracy & Technology*.

Bradshaw, S.; and Howard, P. 2019. The global disinformation order: 2019 global inventory of organised social media manipulation. *Communication Technology and New Media Commons*.

Broderick, R. 2019. Here's What The Mueller Report Says About Russian Trolls Using Social Media To Tamper With The 2016 Election. *BuzzFeed News*.

Cima, L.; Mannocci, L.; Avvenuti, M.; Tesconi, M.; and Cresci, S. 2024. Coordinated behavior in information operations on Twitter. *IEEE Access*.

Cordey, S. 2019. Cyber Influence Operations: An Overview and Comparative Analysis. Technical report, ETH Zurich.

Cresci, S. 2020. A decade of social bot detection. *Communications of the ACM*, 63(10): 72–83.

Cresci, S.; Di Pietro, R.; Petrocchi, M.; Spognardi, A.; and Tesconi, M. 2016. DNA-inspired online behavioral modeling and its application to spambot detection. *IEEE Intelligent Systems*, 31(5): 58–64.

Cresci, S.; Di Pietro, R.; Petrocchi, M.; Spognardi, A.; and Tesconi, M. 2017. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In *Proceedings of the 26th international conference on world wide web companion*, 963–972.

Cresci, S.; Di Pietro, R.; Spognardi, A.; Tesconi, M.; and Petrocchi, M. 2023. Demystifying Misconceptions in Social Bots Research. *arXiv*.

Cryst, E.; Robertson, R. E.; Schechter, N.; and Thiel, D. 2021. Maduro's Menagerie: Twitter Client Usage in a

Venezuelan Twitter Operation. *Stanford Internet Observatory*.

Davis, C. A.; Varol, O.; Ferrara, E.; Flammini, A.; and Menczer, F. 2016. BotOrNot: A System to Evaluate Social Bots. In *Proceedings of the 25th International Conference Companion on World Wide Web*, 273–274.

DiResta, R.; Grossman, S.; and Miller, C. 2019. Analysis of Twitter Takedown of State-Backed Operation Attributed to Saudi Arabian Digital Marketing Firm Smaat. *U.S. Senate Documents*.

DiResta, R.; Shaffer, K.; Ruppel, B.; Sullivan, D.; Matney, R.; Fox, R.; Albright, J.; and Johnson, B. 2019. The tactics & tropes of the Internet Research Agency. *U.S. Senate Documents*.

Gallagher, S. 2019. Rent-a-troll: Researchers pit disinformation farmers against each other. *Ars Technica*.

Gorwa, R.; and Guilbeault, D. 2020. Unpacking the Social Media Bot: A Typology to Guide Research and Policy. *Policy & Internet*, 12(2): 225–248.

Grimme, C.; Preuss, M.; Adam, L.; and Trautmann, H. 2017. Social Bots: Human-Like by Means of Human Control? *Big Data*, 5(4): 279–293.

Hegelich, S.; and Janetzko, D. 2016. Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet. In *Proceedings of the international AAAI conference on web and social media*, volume 10, 579–582.

Ingram, D. 2017. Facebook says 126 million Americans may have seen Russia-linked political posts. *Reuters*.

Knight, W. 2023. It Costs Just $400 to Build an AI Disinformation Machine. *Wired*.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, 591–600.

Lister, T.; Sciutto, J.; and Ilyushina, M. 2017. Putin's 'chef,' the man behind the troll factory. *CNN*.

Luceri, L.; Pantè, V.; Burghardt, K.; and Ferrara, E. 2024. Unmasking the web of deceit: Uncovering coordinated activity to expose information operations on twitter. In *Proceedings of the ACM on Web Conference 2024*, 2530–2541.

Mazza, M.; Cresci, S.; Avvenuti, M.; Quattrociocchi, W.; and Tesconi, M. 2019. Rtbust: Exploiting temporal patterns for botnet detection on twitter. In *Proceedings of the 10th ACM conference on web science*, 183–192.

Neudert, L.-M.; Kollanyi, B.; and Howard, P. N. 2017. Junk News and Bots during the German Federal Presidency Election: What Were German Voters Sharing Over Twitter? Technical report, Univeristy of Oxford.

Polychronis, C.; and Kogan, M. 2023. Working Together (to Undermine Democratic Institutions): Challenging the Social Bot Paradigm in SSIO Research. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2): 242:1–242:30.

Ruths, D. 2019. The misinformation machine. *Science*, 363(6425): 348–348.

Seddon, M. 2014. Documents Show How Russia's Troll Army Hit America. *BuzzFeed News*.

Skrondal, A.; and Rabe-Hesketh, S. 2003. Generalized linear mixed models for nominal data. In *Proceedings of the American Statistical Association*, 3931–3936.

Starbird, K.; Arif, A.; Wilson, T.; Van Koevering, K.; Yefimova, K.; and Scarnecchia, D. 2018. Ecosystem or echo-system? Exploring content sharing across alternative media domains. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Theohary, C. A. 2020. Defense primer: information operations. Technical report, Library of Congress.

Weick, K. E. 1993. The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. *Administrative Science Quarterly*, 38(4): 628–652.

Yang, D.; Halfaker, A.; Kraut, R.; and Hovy, E. 2016. Who Did What: Editor Role Identification in Wikipedia. *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1): 446–455.

Ethics Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, answering this research question would advance SSIO research (as mentioned in the Introduction and Discussions) without exacerbating divides or disrespecting cultures, as we talk about SSIOs in aggregate without commenting on the morality of the specific actors or sponsors involved. As mentioned in the Data subsection under Methods, we also utilize anonymized data to better protect user privacy.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, the claims made in our abstract and introduction are directly drawn from our Findings and Discussion sections. We also clearly outline the scope of our analysis as engaging with the SSIO context, as opposed to other forms of misinformation propagation.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, in our Methods section, we provide justification for using embedded Digital-DNA and GLMMs, as well as provide literature support for our decisions regarding data encoding and operationalization.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, see the Data subsection under Methods.
   (e) Did you describe the limitations of your work? Yes, see the Limitations and Future Work subsection under Discussion.
   (f) Did you discuss any potential negative societal impacts of your work? No, we do not discuss the main potential negative societal impact of our work – a guide on how to build better operations, as this would

require optimization considerations that we do not include in our analysis here. More broadly, the potential negative impacts are in reality quite low, as we report on the state of operations conducted by state actors that have progressed well beyond the means of the operations analyzed here (see Discussion for further details).

(g) Did you discuss any potential misuse of your work? No, the main potential for misuse of our work comes from state actors wanting to weaponize our findings on SSIO effectiveness. We describe how this is unlikely above.

(h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, we document research design choices and include material in the appendix in order to aid in reproducebility. We also work with data that is anonymized in order to protext individuals who were engaged (but not directly involved) with SSIO activity.

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, see this checklist.

2. Additionally, if your study involves hypotheses testing...

(a) Did you clearly state the assumptions underlying all theoretical results? NA

(b) Have you provided justifications for all theoretical results? NA

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA

(e) Did you address potential biases or limitations in your theoretical framework? NA

(f) Have you related your theoretical results to the existing literature in social science? NA

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA

(b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? Yes, the Information Operations archive is explicitly mentioned, with a link provided to this source in the footnotes.

(b) Did you mention the license of the assets? No, we did not mention the license of this dataset, as it is not directly provided by X. Further information about the dataset and terms of use, however, are available at the link mentioned above.

(c) Did you include any new assets in the supplemental material or as a URL? NA

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? No, we did not discuss user consent as the mechanism for this process is available at the archive website, which we direct readers to.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, we discuss how PII was anonymized by the X Team, and how SSIOs engage in messaging that targets divisive issues, which will likely be offensive.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? NA

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)? NA

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA

(d) Did you discuss how data is stored, shared, and deidentified? NA

Appendix

Term	Engagement		Follower Count		Longevity	
	Effect	Sig,	Effect	Sig.	Effect	Sig.
Agent Type:Automated	0.87	****	0.66	****	0.81	****
Agent Activity:Amplification	0.18	****	0.33	****	0.60	****
Activity Frequency:Frequent	0.64	****	0.76	****	0.71	****
Focus:SF	1.29	****	1.35	****	1.22	****
Focus:DF	4.13	****	1.68	****	1.14	***
Focus:NF	0.02	*	0.14		0.00	
Inflow from Bots (log)	1.31	****	1.28	****	1.03	****
Outflow to Bots (log)	0.98	**	0.90	****	1.00	
Inflow from Humans (log)	2.08	****	1.33	****	1.07	****
Outflow to Humans (log)	0.85	****	0.85	****	0.95	****
Outflow to Real Users (log)	1.50	****	1.48	****	1.38	****
Automation Usage (%)	1.01		1.00		0.99	*
Operation Size (log)	0.61		0.70		0.82	
Network Density	0.83		0.86		0.91	

Table 2: Effects and statistical significance for Zero-Inflated Negative Binomial regressions on Engagement, Follower Count, and Longevity. Significance is coded with $*$: $p > 0.5$, $p > 0.01 : *$, $p > 0.001 : **$, $p > 0.0001 : ***$, $p < 0.0001 : ****$. In general, we see that agent-level terms have significant impact on agent success, while operation-level ones do not.

Term	Engagement		Follower Count		Longevity	
	Effect	Sig.	Effect	Sig.	Effect	Sig
Agent Type:Automated	1.14	****	0.86	****	0.89	****
Inflow from Bots (log)	1.31	****	1.35	****	1.06	****
Outflow to Bots (log)	1.00		0.91	****	0.99	
Inflow from Humans (log)	2.32	****	1.42	****	1.11	****
Outflow to Humans (log)	0.71	****	0.81	****	0.91	****
Outflow to Real Users (log)	1.39	****	1.35	****	1.32	****
Automation Usage (%)	1.00		0.99		0.98	*
Operation Size (log)	0.52	*	0.64	*	0.79	*
Network Density	0.76		0.82		0.89	

Table 3: Effects and statistical significance for Zero-Inflated Negative Binomial regressions on Engagement, Follower Count, and Longevity when we do *not* account for agent role. Significance is coded with $*$: $p > 0.5$, $p > 0.01 : *$, $p > 0.001 : **$, $p > 0.0001 : ***$, $p < 0.0001 : ****$. We see that in comparison to when we do account for role, automated accounts are now not strictly worse than humans. We also see that operation size becomes a significant predictor of agent success, but no other operation-level terms do.