Philosophy & Technology (2023) 36:35
https://doi.org/10.1007/s13347-023-00640-9

RESEARCH ARTICLE

Liars and Trolls and Bots Online: The Problem of Fake Persons

Keith Raymond Harris[1]

Received: 12 January 2023 / Accepted: 25 April 2023

Abstract
This paper describes the ways in which trolls and bots impede the acquisition of knowledge online. I distinguish between three ways in which trolls and bots can impede knowledge acquisition, namely, by deceiving, by encouraging misplaced skepticism, and by interfering with the acquisition of warrant concerning persons and content encountered online. I argue that these threats are difficult to resist simultaneously. I argue, further, that the threat that trolls and bots pose to knowledge acquisition goes beyond the mere threat of online misinformation, or the more familiar threat posed by liars offline. Trolls and bots are, in effect, fake persons. Consequently, trolls and bots can systemically interfere with knowledge acquisition by manipulating the signals whereby individuals acquire knowledge from one another online. I conclude with a brief discussion of some possible remedies for the problem of fake persons.

Keywords Bots · Misinformation · Social epistemology · Social media · Testimony · Trolls

1 Introduction

A 2018 indictment by the U.S. Justice Department lays out a series of allegations against members of Russia's Internet Research Agency (IRA). Members of the IRA are alleged to have taken various measures to influence the 2016 U.S. Presidential Election and the political system more generally. According to one such allegation, members of the IRA:

> [C]reated hundreds of social media accounts and used them to develop certain fictitious U.S. personas into "leader[s] of public opinion" in the United States. (U.S. District Court, 2018)

✉ Keith Raymond Harris
keithraymondharris@gmail.com

[1] Ruhr-Universität Bochum, Bochum, Germany

Published online: 03 May 2023

Among other purposes, members of the IRA are alleged to have used these fake American personas to shape public opinion and to organize political events. In one noteworthy instance, Facebook groups established by the IRA organized two competing protests at the Da'Wah Islamic Center in Houston, Texas. The event was one of the hundreds organized by the IRA (Cosentino, 2020: ch. 2) and vividly illustrates the potential of inauthentic online activity to spill over into the offline world.

The IRA's activities help to make salient the possibility that various forms of misinformation, spread principally online, might be used to influence public opinion on a large scale. While the actual effectiveness of misinformation campaigns in shaping public opinion has been questioned (Mercier, 2020; Eady et al., 2023), such campaigns remain an important topic for epistemological study. Indeed, even if the impact of misinformation on public opinion is largely overstated, this point would itself be worthy of epistemological consideration. Moreover, even if misinformation campaigns are not enough to sway elections, such campaigns can—as the competing protests mentioned above illustrate—manipulate individual activities on a smaller scale.

This paper joins a growing body of work on the epistemology of misinformation and social media. Whereas previous epistemological work has tended to focus on fake news (Bernecker et al., 2021; Fallis & Mathieson, 2019; Gelfert, 2018; Grundmann, 2020; Jaster & Lanius, 2018; Levy, 2017; Rini, 2017), deepfakes (Fallis, 2021; Harris, 2021; Rini, 2020), and other forms of epistemically deficient content, the present study focuses on some of the entities responsible for spreading such content, namely, bots and trolls. These are entities to which epistemologists have thus far devoted surprisingly little attention[1]. In what follows, I will discuss the ways in which trolls and bots can in principle undermine the acquisition of knowledge on social media and in other online spaces. I will argue that trolls and bots, understood as fake persons, undermine the acquisition of knowledge by interfering with one or more of the truth, belief, and warrant conditions on knowledge. A key upshot of the discussion to follow is that trolls and bots are not simply conduits for the spread of fake content. Rather, being fakes themselves, the threats that trolls and bots pose to knowledge go beyond the simple transmission of misinformative content. I conclude with some brief remarks on possible remedies for the threats posed by fake persons online.

2 Background: Fake Persons and Threats to Knowledge

This paper is principally concerned with two types of entities encountered on online social networks and in other online spaces. While trolls and bots differ in important respects, members of both categories can and sometimes do operate under fabricated identities that are presented as real identities. Trolls and bots that present themselves in these ways can be regarded as fake persons. Notably, those who adopt online personas that differ from their own need not be fake persons, on the present understanding. One might, for example, post under a pseudonym that is not

[1] But see Jennifer Lackey (2021) and Regina Rini (2021) for some rare exceptions.

presented as belonging to a real person, without thereby acting as a fake person. In what follows, I will be concerned with the consequences of interactions with fake persons for the spread of knowledge online. The epistemic problem of fake persons is not entirely new. The construction of false identities has, for instance, long been an element of spycraft. But with the rise of trolls and bots, encountering human and non-human entities operating under false identities has become commonplace. This phenomenon calls for epistemological consideration, especially in light of our epistemic dependence on others (Hardwig, 1985).

A productive discussion of trolls and their effects requires a disambiguation between some related phenomena. Until recently, trolls were mostly understood as individuals who exhibit provocative behavior online, often seeking to promote conflict within online groups for the sake of their own amusement (Fichman & Sanfilippo, 2016; Hardaker, 2010: 237). Key to this understanding is that trolls regard trolling chiefly as a way of having fun online (Krappitz, 2012; Shachaf & Hara, 2010).

While uses of the terms "troll" and "trolling" to describe politically and ideologically motivated individuals and behaviors are not new, it is only recently that these usages have come to the fore. In recent years, commentators have described organized networks of individuals pushing propaganda as trolls and have used the terms "troll farm" (Morrison, 2021) and "troll factory" (Linvell & Warren, 2020) to describe these networks themselves. For example, Russian troll factories have been implicated in interventions in a range of geopolitical contexts, including the 2016 Brexit Referendum (Bastos & Farkas, 2019) and the 2016 US Presidential election[2] (Linvell & Warren, 2020; US District Court, 2018). State-sponsored trolls belong to the wider category of politically or ideologically motivated trolls, whose epistemic effects will be of primary interest here. Such trolls have in common with trolls in general the tendency to post insincerely online but differ in their motivations and, consequently, in some of their behaviors. For example, whereas trolling may be performed sporadically and under one's own name, politically and ideologically motivated trolls, especially state-sponsored ones, tend to engage in long-term influence campaigns carried out under fabricated identities. Given this distinctive characteristic, it is not to be expected that all remarks on the epistemic consequences of trolls that take on false identities are applicable to trolls more widely. Thus, in what follows, I focus on this narrower category of trolls and use the term "troll" to denote these.

Trolls operate in a range of spaces online, including social media platforms (Bastos & Farkas, 2019; Golovchenko et al., 2020) and comment sections (Chen, 2015;

[2] The actual efficacy of trolls and bots in shaping political outcomes is a matter of contention. For example, while the Russian use of troll factories to influence the 2016 US Presidential election has received an enormous amount of media attention, a major recent study suggests that the influence of Russian Twitter trolls on American attitudes in the lead up to the election was trivial or non-existent (Eady et al., 2023). More generally, the question of whether trolls and bots have thus far been successful in interfering with knowledge is an empirical one and cannot be resolved here. Instead, my aim in this paper is to consider how trolls and bots might, in principle, interfere with knowledge. Notably, as we will see, one mechanism by which trolls and bots might do so is through exaggeration of their impacts on online epistemic environments.

Knustad, 2020; Morrison, 2021). On social media, trolls may engage in any of the actions available to ordinary users, including creating original posts, replying, sharing, liking, and following. By these mechanisms, trolls influence the spread and reception of information online. I begin to explore the consequences of this point in Sect. 3.

Bots are computer algorithms that can emulate the online behavior of human persons. For instance, bots that are deployed on social media platforms—social bots—can typically post original content, reply, share, like, and follow other accounts (Daniel & Millimaggi, 2020). Like trolls, bots can also operate outside of social media, including in the comments sections of news outlets and video-sharing platforms. Bots may be deployed for a range of purposes, some of which are epistemically neutral or even straightforwardly beneficial. For example, it is increasingly common for commercial companies to utilize bots for the purposes of advertising and responding to consumer queries. At least so long as these bots are clearly identified as such, they can streamline the acquisition of information without causing undue confusion. It is worth noting, however, that commercial bots that are not clearly labeled as such might give rise to many of the problems raised below[3].

Some bots are put to nefarious ends. Like trolls, bots can be deployed in an organized fashion to shape online narratives. Bots have often been implicated in the spread of misinformation online (Bastos & Mercea, 2018, 2019; Broniatowski et al., 2018; Ferrara, 2020), and are often driven by state actors (Alsmadi & O'Brien, 2020; Prier, 2017). In other cases, individual actors may employ bots to amplify their online presences (Stella et al., 2019). Overall, bots are responsible for a significant portion of the posts on social media (Ferrara et al., 2016; Wojcik et al., 2018), especially those surrounding major political events (Bastos & Mercea, 2018, 2019; Caldarelli et al., 2020; Jones, 2019). In this way, bots may exert substantial influence on the social epistemic environment.

While trolls and bots can take many forms online, those that take on false human identities can sensibly be regarded as fake persons, and it is these entities with which I will be concerned in what follows. The line between trolls and bots is sometimes blurry, as in the case of "cyborg" accounts, the behavior of which is partially human-controlled and partially automated (Chu et al., 2012). Some users might use cyborg accounts simply to enhance their engagement online, while still taking responsibility for the content automatically posted. However, trolls disguised as real persons might also use partially automated accounts to extend their influence. Such cyborgs can likewise be categorized as fake persons, but because cyborgs essentially combine features of trolls and bots, I do not discuss cyborgs independently in what

[3] Relatedly, even if bots are clearly labeled as such, the manner of their communication might lead to confusion as to their capabilities. Thus, for example, the impressively fluid communication of recently developed chatbots has been mistaken for evidence of consciousness (Tiku, 2022). Additionally, it has been argued that the use of emojis by chatbots should be avoided, as the use of emojis might be deceptive with respect to the emotional capacities of the chatbots in question (Véliz, 2023). While there are many concerns to be raised concerning the potential for confusion about the capacities of even bots explicitly identified as such, I focus here on concerns that arise from bots masquerading as persons.

follows. Rather, I will discuss the epistemic effects of trolls and bots, construed as fake persons.

It might be thought that contrasting trolls and bots with "real persons" reflects an oversimplified binary. As an anonymous referee has rightly pointed out, the activities of biological humans—especially within social media contexts—are routinely, in some sense, inauthentic. Consider a few examples. Ordinary social media users regularly present idealized depictions of their lives, their appearances, and their abilities. Such depictions sometimes cause non-trivial harms. For example, the exposure to manipulated photos on Instagram has been shown to cause adolescent girls to develop unrealistic standards of female beauty and, consequently, anxieties about their own bodies (Kleemans et al., 2018). As a second example, consider that individuals sometimes pretend to hold certain political attitudes in order to signal their loyalties within experimental settings (Hannon & de Ridder, 2021; Levy, 2022: ch. 1; Ross & Levy, 2023; Schaffner & Luks, 2018). Given the significant rewards of such signaling on social media, it is highly plausible that individuals sometimes engage inauthentically with online political content for the sake of political advantage. In short, there is good reason to suspect that even real persons often act fake online. Consequently, I suspect that some of the problems I will suggest are caused by trolls and bots and are also caused by the inauthentic activities of real persons. However, I focus here on trolls and bots for two reasons. First, the banal fakery in which ordinary individuals engage is arguably relatively benign precisely because it is so ubiquitous. Because nearly all of us present somewhat idealized versions of ourselves online and sometimes act online in such a way as to express our loyalties, we have a reasonably good sense of how to interpret the depictions of others' lives that we encounter online. The caveat to this first point is that ordinary individuals engage in banal fakery to different degrees, in different contexts, and for different reasons. This brings us to a second reason for focusing narrowly on trolls and bots. By first attending to these relatively pure instances of fakery, we can establish insights that might then be applied to the more mixed, inauthentic activity.

To conclude this introduction to trolls and bots, I wish to emphasize three points. First, I am concerned here with trolls and bots construed as fake persons, and so only with those trolls and bots that post under fabricated human identities. Not all trolls and bots are fake persons in this sense. Some of the claims made in what follows about bots and trolls construed as fake persons apply to bots and trolls that do not operate under fake human identities. However, concentrating on the narrower groups of bots and trolls that operate under fabricated human identities will help to focus the discussion. In what follows, I refer to these relatively narrow groups simply as bots and trolls, or as fake persons, for the purpose of simplicity. Second, I do not mean to imply here that bots and trolls raise precisely the same epistemic concerns. For example, the relative ease of coordinating large numbers of bots might raise distinctive epistemic concerns, while the human intelligence of trolls might raise others. I focus here on the shared epistemic consequences of these entities, in order to emphasize how fake persons, which may exist in distinct forms, generally impact others' prospects for acquiring knowledge online. Finally, I am principally concerned in this paper with the impacts of bots and trolls at the time of this writing. Future technological developments related to the artificial intelligence of bots and

the detection of bots and trolls will likely alter the severity of the consequences of such fake persons. However, part of what I want to illustrate here is that even relatively rudimentary bots have significant consequences for the spread of knowledge online.

Throughout this paper, I will assume that knowledge is warranted by true belief, where warrant is simply an epistemic condition or set of conditions that bridges the gap between true belief and knowledge. Warrant might thus be understood in terms of internalist or externalist justification, a modal condition, or some combination of these. Given this approach to knowledge, there are at least three pathways by which fake persons might compromise knowledge. Fake persons might interfere with the truth condition, the belief condition, or the warrant condition on knowledge. Insofar as fake persons threaten each of these conditions, fake persons constitute what I will call a deceptive threat, a skeptical threat, and an epistemic threat[4], respectively. In Sects. 3–5, I discuss the ways in which fake persons constitute each of these kinds of threats.

3 The Deceptive Threat of Fake Persons

In this section, I discuss the ways in which fake persons are likely to cause false beliefs in at least some audiences. A core aim of this section is to show that the deceptive threat of fake persons is not simply the familiar deceptive threat of liars carried out online. Liars are, in effect, conduits for information believed to be false. While bots and trolls also sometimes pass on false or misleading content, the deceptive threat of such fake persons is not limited to this communicative role. The comparatively deep deceptiveness of trolls and bots is rooted in their fake identities.

Fake persons constitute deceptive threats to the extent that they are likely to cause false beliefs. As I suggested above, the deceptive threat of fake persons is, in part, linked to the deceptive threat of other sorts of fakes. For example, to the extent that trolls and bots distribute fake news and fake science, such fake persons thereby amplify the deceptive threat of these other fakes. In fact, the deceptive activities of fake persons—especially bots—often take the form of simply passing on—by way of Retweets or similar sharing functions—fake news (Ferrara, 2020; Jones, 2019). Passing along existing content in this way is a simpler task than introducing original deceptive content. Still, trolls and bots likewise engage in the latter, more sophisticated form of deception.

That fake persons can spread fake content—whether original or recycled—does not by itself show that such fake persons constitute a deceptive threat. If trolls and bots were especially unconvincing fake persons, those exposed to their content might place little weight on it and thereby avoid deception. However, trolls and bots

[4] It bears noting here that, in the context of a discussion of the epistemology of deepfakes, Don Fallis (2021) uses the term "epistemic threat," in a more encompassing way than the one adopted here. Fallis effectively treats what I call deceptive threats, skeptical threats, and epistemic threats as "epistemic threats" in his sense. In a discussion of online misinformation, Keith R. Harris (2022) follows Fallis in using "epistemic threat" in this relatively encompassing way.

pose a deceptive threat by way of the content they spread by, in part, deceiving with regard to their identities. In short, some of the false beliefs fake persons are prone to causing have to do with the identities of those very fake persons. Such deception can be facilitated by various means. For example, some trolls use pictures taken from elsewhere on the internet (Chen, 2015; Hampton, 2019), including others' social media accounts (O'Sullivan, 2017). Likewise, bot accounts often utilize stolen photos of real humans (Ferrara et al., 2016: 100).

Trolls and bots are not merely potentially deceptive with regard to their identities; empirical studies suggest that such fake persons succeed in deceiving ordinary human users. Early empirical studies indicate that bots on social media are effective at passing themselves off as real people, with neither human users (Cresci et al., 2017) nor bot-detection algorithms (Martini et al., 2021) being able to reliably identify bots as such. The detection of trolls on social media has likewise proven difficult, although various methods have been proposed (e.g., Fornacciari et al., 2018). Notably, even if sophisticated technological or non-technological methods for identifying bots and trolls are developed, the mere existence of such methods will not by itself help ordinary users to distinguish between real and fake persons online. At a minimum, ordinary users will require ready access to this information.

Given that fake persons are at least somewhat effective at deceiving others into thinking that they are real persons, there is reason to expect that fake persons can further deceive audiences by way of the content they share. In sharing content like fake news and fake science, trolls and bots not only expose audiences to that content, but also provide evidence that real people believe that content. In this way, fake persons spread not only fake first-order evidence, but also fake high-order evidence—evidence of the quality of first-order evidence. Suppose, for example, that a troll or bot shares a fake news story alleging the involvement of a prominent politician in a corruption scheme. In addition to the allegation and whatever evidence is contained in the story, the fake person thereby provides some higher-order evidence as to the warrant for the allegation. Notably, the deceptive higher-order evidence furnished by a single troll or bot sharing a fake story might well be negligible. However, coordinated bot and troll campaigns can give the appearance that bogus content is widely deemed to have merit. For example, bots and trolls on Twitter often participate in vast "Retweet networks," which are characterized by the mass re-sharing of content. The re-sharing of content through such networks serves not only to distribute it but, at the same time, to give that content the appearance that it has been widely considered and deemed worthy of sharing. While the norms of social media testimony are unsettled—such that the re-sharing of content cannot straightforwardly be interpreted as an endorsement of that content (Marsili, 2021; Rini, 2017)—some human users exposed to widely-reshared content will no doubt interpret the apparent popularity of the content as evidence of its legitimacy and will consequently be deceived (Ferrara et al., 2016: 98-99).

The tools by which troll and bot campaigns can distort perceptions of public attitudes toward online content are not restricted to the use of shares. The ability to "like" content, which is available on social media and a broad range of other online platforms, can shape perceptions of public attitudes toward liked content, thereby artificially inflating the credibility of that content. While philosophers have

argued that liking something online does not straightforwardly signify endorsement (McDonald, 2021), empirical results suggest that likes are often interpreted as "endorsement cues," which signal the credibility of the content liked (Luo et al., 2022). Summarizing early work on online credibility, Johan Jessen and Anker Helms Jorgensen (2012) write that likes, shares, and other forms of social validation figure significantly in judgments of credibility for information retrieved online. By impersonating real persons, trolls and bots can exploit this tendency to rely on social indicators of credibility.

Notably, a fake person might distort perceptions not only of public opinion in general but of the opinions of some segments of the population. For example, white supremacists have sometimes impersonated black Twitter users with the aim of distorting perceptions as to the views of this segment of the population (Rashid, 2017). Such efforts are especially pernicious insofar as they play on and encourage existing stereotypes among social media users. Some such campaigns have been combated through the coordinated investigative efforts of genuine black feminists (Hampton, 2019).

The concern about fake persons deceiving by way of the content they share or otherwise promote might seem too quick, unless there is good reason to expect that fake persons are especially likely to share false or misleading content. It is not an essential feature of the category of fake persons that they usually or always act so as to deceive audiences about more than their identities. It might thus fairly be asked why mistaking trolls and bots for real persons exposes one to further deception. There are at least two reasons for this. First, as the preceding paragraphs anticipate, deception by fake persons need not involve the sharing of fake content. Consider an example. Bot and troll accounts might share accurate content or content that is not subject to assessment in terms of accuracy[5], in a way that is nonetheless deceptive insofar as the mass-sharing of the content in question presents a misleading picture of public opinion toward it. In short, fake persons might be deceptive with respect to what people think, even if they are not deceptive with respect to the content of the material shared.

Second, there is good reason to expect fake persons to share false or misleading content relatively frequently. Admittedly, it should be acknowledged that disguising or otherwise concealing one's identity has some legitimate applications, and need not indicate the unreliability of the subject's testimony in all cases. For example, testifiers in vulnerable positions may have strong reasons to conceal their identities in certain contexts. However, protection of the vulnerable does not, in typical cases, require the fabrication of a distinct identity. Moreover, there is cause for skepticism as to the reliability of information received from trolls and bots. This includes information that is introduced by the troll and bot accounts, as well as information merely passed on by such accounts through sharing, retweeting, and related methods. I will subsequently use the general term "posts" to capture all such contributions. First, the fabrication of an identity plausibly removes whatever default basis there is for trust in the information a subject provides. Typically, testifiers have a practical incentive

[5] For an example of the latter, consider that mass-sharing of a photo of a political candidate, X, together with the caption "X for President," might distort perceptions of that candidate's degree of support.

to adhere to the truth, as departures from the truth are likely to damage their reputations and, especially, their perceived credibility. The extent to which this point generalizes to sharing is an open question for social epistemologists, largely because the sharer's attitude toward the shared content is often unclear (Marin, 2021; Marsili, 2021; Rini, 2017). However, even if the sharing of bad information does not redound upon the sharer in the same way that bad testimony redounds upon the testifier, there remains an incentive for careful sharing practices. Habitually sharing bad information can be expected to cause one to develop a reputation for doing so and perhaps for having poor judgment. In general, then, individuals typically have reason to introduce and to share information that they at least perceive to be accurate. But this incentive does not apply to trolls and bots—or, more conservatively, does not apply to the same degree. By disguising one's identity online—either by posting under a fabricated identity or deploying bots—one can substantially reduce the practical incentive to avoid posting bad information. So long as the disguise is effective, bad epistemic practices will not redound on the individual.

It might be objected that there is an incentive for even trolls and bots to typically post truthfully. Bad posting practices will not damage the reputation of the human individuals behind troll and bot accounts, but they will plausibly reduce the perceived credibility of the troll and bot accounts themselves. Such accounts will be most effective if they are not immediately dismissible as non-credible. For example, if a state actor aims to shift public opinion on a certain issue, this goal might be best achieved through posts by trolls and bots that are perceived as credible. The dissemination of accurate information to establish credibility for future deceptive efforts is a familiar strategy sometimes called "pre-propaganda" (Ellul, 1973; Golovchenko et al., 2020). Given the importance of establishing credibility, one might argue that, even for trolls and bots driven by the ultimate aim of deception, there is an incentive to typically post truthfully.

Three responses to this objection are in order. First, it is consistent with the objection that mistakenly perceiving trolls and bots as real persons contributes to deception about certain key points. For example, even if trolls and bots deployed by a state actor mostly post to establish credibility with respect to a target issue, mistaking these trolls and bots for real persons will facilitate one's deception with respect to that target issue. In principle, this means that even fake persons deployed to deceive about a particular issue might have a net positive epistemic impact, supposing their deception is outweighed by the earlier spread of truths[6]. However, a second response to the present objection is more pessimistic as to the overall epistemic impacts of fake persons. It is not clear that the value of high perceived credibility provides an incentive to post truthfully in most cases. Perceived credibility requires at most that one's posts are perceived as true by the audience, not that they are in fact true. If a given audience already has an inaccurate picture of reality, trolls, and bots can maintain and even heighten their perceived credibility with that audience through inaccurate posts. Moreover, it would be a mistake to suppose that an account's perceived credibility is owed primarily to its posting history. While attention to track records

[6] Thanks to an anonymous referee for raising this point.

plausibly contributes to the perceived credibility of others in more traditional relationships and perhaps in long-term online interactions, many online platforms expose users to posts from accounts with which they are not familiar. The epistemic track records of such accounts can be assessed, if at all, only with substantial effort (cf. Rini, 2021: 40). Consequently, rather than assessing the perceived credibility of online accounts according to their track records, users are more likely to rely on crude strategies like consulting the account's biography, profile picture, and follower count. As we have seen, trolls and bots can fabricate biographical details to signal credibility. By acting in concert, such fake persons can likewise achieve misleading follower counts. Finally, even if the perceived credibility of a troll or bot account is damaged by a poor track record, identities can be shed relatively easily online—limiting the costs of low perceived credibility. For example, trolls may open new accounts under which to post or may change the biographical details associated with their accounts. Likewise, those employing bots may simply employ new bots or may instead allow existing bots to take on new identities.

I have thus far focused on the deception likely to befall thoroughly credulous audiences—those inclined to accept trolls and bots as real persons and to believe the content such fake persons share. But some users who interpret re-sharing as an endorsement and who mistake fake persons for real ones are likely to be deceived in a final, further way. To get an initial grasp on this further concern, consider the following case. Suppose that a given internet user, Tim, is entirely unaware of the existence of trolls and bots, but nonetheless, encounters them regularly online. Knowing nothing of trolls and bots, Tim can only think of the behavior of these entities as the behavior of real human beings. Tim will likely come to believe, rationally in light of the evidence, that real human persons are often entirely unreasonable and unreliable testifiers, at least in the online environment. In short, Tim might come to be deceived concerning what ordinary human beings are like[7].

The present case is contrived, but not unrealistic. Individuals no doubt have varying degrees of awareness of trolls and bots, and especially of the degree of effort that is put into making trolls and bots resemble genuine human persons. Consequently, it is likely that some individuals routinely mistake trolls and bots for real persons. Such mistakes are likely to lead individuals to, again rationally, reduce their estimations of the reliability of online human testifiers, and perhaps human testifiers more generally. Mistaking online trolls and bots for real persons often enough will lead one to be skeptical of the content one encounters online. I discuss some further consequences of this point in Sect. 4. For the present, it bears emphasizing that the reduced estimation of real persons that is likely to result from misidentifying trolls and bots is not solely an epistemic matter. Tim, and those like him, will likely

[7] It is worth acknowledging here that phenomena other than fake persons can cause similar problems. For example, out-of-context clips or simply non-representative examples are sometimes used to frame members of the political opposition as especially unreasonable. Still, fake persons add a new dimension to this issue as they are often encountered directly and, hence, may appear to better represent what persons are really like.

mistakenly ascribe the negative traits that are actually exemplified by fake persons to real ones.

It might be objected that mistaking trolls and bots for real persons does not happen frequently enough to significantly alter perceptions of real persons. Three responses to this objection are in order. First, the objection may well underappreciate the prominence of trolls and bots online. Trolls and bots are responsible for a great deal of social media posts, especially around certain topics (Bessi & Ferrara, 2016; Ferrara, 2020; Morrison, 2021; Stukal et al., 2022). Second, while I am chiefly concerned here with the threat currently posed by fake persons, it is worth noting that this threat may increase as the utility of troll and bot campaigns is recognized and as automation technology advances.

A third response is that trolls and bots might dramatically alter perceptions of real persons, even if this effect is not entirely general. As I have suggested above, the perceived credibility of troll and bot accounts may be facilitated by the fabrication of biographical details. The suggestion here is that the inclusion of certain biographical details may increase the perceived credibility of an account's posts. Likewise, routinely posting noncredible content may reflect poorly not only on an individual account, but also on accounts with similar biographical details. For example, if trolls and bots that identify themselves as supporters of the Green Party consistently post misinformation, this activity will likely undermine the perceived credibility of real human users who identify themselves as members of the Green Party. More generally, for those who mistake trolls and bots for real persons, the negative behaviors of these fake persons encourage false beliefs critical of real persons with biographical details similar to those adopted by the trolls and bots in question.

4 The Skeptical Threat of Fake Persons

Whereas deceptive threats challenge the truth condition on knowledge, skeptical threats challenge the belief condition. Thus, as presently understood, skepticism is a psychological phenomenon. The close connection between deception and skepticism is central to epistemology, even if it is not always considered in terms of threats. For example, Descartes treats his history of being taken in by false theories as a basis for resisting belief in any subsequent theory that cannot be placed on an unimpeachable foundation. More generally, Descartes uses the possibility of deception by a malevolent and powerful being as a means of resisting the psychological pull of belief. While the abandonment of belief is a desirable intermediary state within Descartes' project, threats to belief are potentially harmful to practical purposes. Thus, I treat the tendency of bots and trolls to discourage certain kinds of beliefs as a skeptical threat. As I argue in what follows, the skeptical threat of fake persons shadows the deceptive threat—for whatever form of deception fake persons might bring about; there is a parallel form of skepticism.

Whereas one unfamiliar with trolls and bots is likely to mistake certain fake persons for real ones, one more familiar with these entities is likely to mistake real persons for trolls or bots, thereby failing to believe that real people are real. Just as fabricated biographical details can be used to encourage the mistaken belief that trolls

and bots are real persons, the recognition that it is possible to fabricate such details encourages some degree of skepticism toward the realness of persons encountered online. For this reason, as we will see in what follows, frequent media reporting on the existence of trolls and bots likely increases the skeptical threat of these fake persons.

The misidentification of real persons as trolls or bots is likely to be especially appealing in some cases. For example, suppose that one finds oneself confronted with a news story that challenges one's political convictions and that has been widely shared online (Koslowska, 2020). Given the knowledge that trolls and bots exist and sometimes share false or misleading information, it will be tempting for one to dismiss both the news story and its apparent popularity as fake. In other words, because trolls and bots sometimes effectively provide fake higher-order evidence, recognition of the existence of trolls and bots may lead one to improperly dismiss as fake the legitimate higher-order evidence furnished by the popularity of a given news story.

More generally, the knowledge that the online environment is populated by bots and trolls enables a convenient strategy for dismissing certain kinds of evidence. Whenever one encounters social evidence—in the form of testimony, shares, likes, and so on—one can dismiss this evidence as the work of trolls and bots. The concern here resembles the concern that Regina Rini (2020) raises for deepfakes, namely, that the existence of sophisticated fake videos in the form of deepfakes makes possible the dismissal of video evidence in general as fake. As Rini emphasizes, such dismissals are likely to be especially attractive when the video evidence in question challenges one's beliefs or interests. In effect, Rini links the deceptive threat of deepfakes to the skeptical threat of such videos. The present point is that the potentially deceptive nature of fake persons encourages skepticism concerning the identities of those one encounters online, of the quality of content shared online, and of apparent endorsements of that content. In this way, the suspicion that a given environment contains fake persons encourages the dismissal of the evidence constituted by content shared online, as well as the higher-order evidence furnished by others sharing or responding positively to that content.

Bots and trolls can encourage skepticism even for those individuals that fail to recognize the existence of such fake persons. As I noted in section 3, those individuals who mistake fake persons for real ones are likely to develop negative attitudes toward certain groups of real persons, and real persons more generally. Some of the negative attitudes in question are likely to be epistemic—concerning, in particular, the credibility of real persons. Insofar as one comes to believe that real persons are non-credible, one will have reason to conclude that the social evidence provided by such persons—in the form of testimony, shares, and likes—is of little evidential import. In this respect, the skeptical consequences of the posts of bots and trolls for those of real persons resemble the skeptical consequences of mistaking the posts of real persons for those of trolls and bots. However, there is an important respect in which mistaking trolls and bots for real persons has more dire consequences. This point is vividly illustrated by returning to the case of white supremacists impersonating black Twitter users. One apparent aim of such campaigns has been to discredit black voices (Hampton, 2019). More generally, mistaking trolls and bots for real

persons may carry over to offline life, leading one to dismiss the evidential significance of what real human persons do in the world at large. As the preceding example illustrates, such activities may be especially pernicious when they target subgroups, especially those already facing imbalances in power.

I have argued in this section that one of the mechanisms by which trolls and bots can interfere with the acquisition of knowledge online is by encouraging skepticism toward even genuine persons and content encountered online. Notably, this suggests that fake persons are likely important tools for certain forms of disinformation—especially those associated with the Russian model—thought to aim at producing doubt rather than false belief (Paul & Matthews, 2016; Pomerantsev, 2014; Rini, 2021). Ironically, some attempts to recognize and confront the threats of fake persons have plausibly enhanced their skeptical threat. Some recent empirical work has suggested, for example, that IRA trolls had limited reach (Eady et al., 2023), interacting mainly with small numbers of highly partisan voters. In contrast, mainstream reporting about IRA trolls had an enormous reach and has plausibly impacted trust in persons and information encountered online. In this way, the entanglement of the deceptive and skeptical threats of fake persons makes it difficult to bring attention to the former without exacerbating the latter.

5 The Epistemic Threat of Fake Persons

I have thus far discussed the deceptive threat and the skeptical threat of fake persons. These threats are alike in that they consist of the distortion of beliefs. Given that what a person believes shapes how that person acts, the deceptive and skeptical threats are of immediate practical concern. In this section, I turn to the epistemic threat of trolls and bots, which consists of the tendency of such fake persons to interfere with the warrant condition on knowledge.

How to substantiate the warrant condition remains a matter of contention among epistemologists. I will not address this matter at any length here, preferring instead to consider some ways in which fake persons can interfere with the warrant condition, regardless of how precisely this is construed. First, the very existence of fake persons degrades the import of certain kinds of evidence. Ordinarily, the biographical details and posting history associated with a given account would likely be interpreted as evidence of the identity of the person associated with that account. For instance, if an account specifies that it belongs to a female attorney located in the UK and has a profile picture and recent posts that accord with these details, this information would naturally be interpreted in favor of concluding that the account is run by such a person. However, in an environment populated by trolls and bots, these forms of evidence carry less weight than they otherwise would. Here, an analogy is helpful. Ordinarily, the outward appearances of animals and plants carry significant information as to the species to which individual organisms belong. However, in environments populated with mimic species resembling the species in question, the markings of individual organisms carry less weight. Previously, Fallis (2021) has used this analogy to discuss the epistemology of deepfakes, and Harris (2022) has applied the analogy to online fakes more generally. Notably, the analogy

is especially fitting in the case of fake persons, which effectively mimic real persons. Just as animal and plant mimics reduce the evidential significance of certain perceptible features of target species, so too do fake persons reduce the evidential significance of biographical details and posting behaviors. Trolls and bots thereby make it more difficult to know who real persons are online, even if the truth and belief conditions are satisfied.

The preceding remarks suggest a template by which fake persons can interfere with the acquisition of knowledge online, especially on social media. We have seen, thus far, that fake persons can deceive by way of posting content—as well as by various forms of social validation including shares and likes, and that fakes and bots can, in this way, encourage skepticism about the legitimacy of posts and social validation. Given the epistemic threat of fake persons, this skepticism may well be reasonable. In other words, trolls and bots might compromise these forms of evidence such that one cannot form warranted beliefs based upon them. In part, the epistemic threat of trolls and bots resembles the epistemic threat of more familiar kinds of liars. Social epistemologists have often suggested that conditions of the background epistemic environment can interfere with the acquisition of warrant from testimony (Adler, 1996; Goldberg, 2007; Graham, 2000; Harman, 1973; Lackey, 2008). For example, suppose that one forms the true belief that *p* is based on sincere and competent testimony that *p* from a speaker, S. Whether or not one thereby comes to know *p* in this way plausibly depends on the conditions of the background epistemic environment. For example, if S is the only sincere and competent speaker in the environment, and one could easily have believed ~*p* based on the testimony of one of the many liars in the environment, one arguably does not come to know *p* through S's testimony. Diagnoses of this case may differ. It may be that one lacks knowledge in this case because one's belief is unsafe (Goldberg, 2007), because S's assertion fails to carry information in the context (Graham, 2000), because the process by which it is formed is unreliable (Schmitt, 2017), because the proximity of liars introduces uneliminated relevant alternatives (cf. Blake-Turner, 2020) or a defeater for one's belief, or some further possibility. Because I aim to illustrate the epistemic consequences of bots and trolls given a wide range of epistemological perspectives, I will not attempt to adjudicate between these diagnoses here.

Instead, the point I wish to emphasize here is that trolls and bots may alter online environments such that they come to resemble, in some respects, a room largely populated by liars, thereby undercutting the power of testimony to transmit knowledge. Consider some examples. After viewing a video online, one may look to the comments section for judgments on the credibility of the video. Ordinarily, the testimony that the video is credible would plausibly provide some evidence to that effect. However, supposing that the comments section is mostly composed of insincere comments from trolls, even the sincere, competent, and correct testimony of a genuine user will carry limited epistemic weight. Consider next the environment on a given social media platform following some major geopolitical event. As one looks to better understand the situation, the ability of sincere and competent testifiers to convey warrants may be undercut by the prevalence of social bots spreading misinformation. Notably, such examples are not farfetched but reflect the actual conditions under which online testimony is often received.

While the epistemic threat of trolls and bots resembles the epistemic threat of liars in some respects, it would be a mistake to conclude the former is just a version of the latter, carried out online. First, as we have seen, trolls and bots are deceptive about their very identities, a dimension of deception atypical of ordinary liars. Second, while some forms of social validation can plausibly be understood as lies—shares and likes, for example—others cannot. For example, follower counts constitute a kind of social validation. That one has a high number of followers on Twitter is some indication of one's epistemic worth, for instance. Yet, the act of following on Twitter cannot plausibly be understood as a lie, even if it serves to artificially boost the credibility of some dishonest figure. Fake persons thus resemble liars in the effects they have on the evidential weight of testimony, but the epistemic threat of fake persons goes beyond that of mere liars.

We form beliefs not only based on the content with which we interact, but also based on how this content is received by others. This point is not specific to the online world. How we react to statements made during a political rally need not be purely a matter of the content of those statements but may be responsive to how these statements are received by other members of the audience. Given the participatory nature of the modern web, the posts and reactions of other users are highly visible. In simple terms, the epistemic threat of fake persons is that such persons can appear and behave in much the same way as real persons online, thereby systematically reducing the value of the signals provided by real persons.

6 Concluding remarks

In the preceding pages, I have discussed the deceptive, skeptical, and epistemic threats of fake persons. As we have seen, these threats are closely interrelated. It is because we recognize the threat of deception by trolls and bots that we are likely to be skeptical of some of the forms of evidence encountered online. This skepticism is well-founded to the extent that, at least in some contexts, trolls and bots genuinely do interfere with the significance of evidence encountered online. The interrelations between these threats make clear the difficulty of resolving them. Without adopting a defensive posture toward evidence encountered in spaces likely populated by trolls and bots, we run the risk of deception. But by adopting such a posture, we risk excessive skepticism.

It might be thought that the problem of fake persons has a straightforward solution. By simply declining to form beliefs on the basis of evidence that might be distorted by trolls and bots, we thereby avoid the problem. However, such a suggestion is misguided for two reasons. First, for all its flaws, social media and other features of the participatory web make possible the rapid spread of information and remove obstacles to epistemic contributions by those who otherwise lack platforms. The suggested solution effectively amounts to succumbing to the skeptical threat with respect to these contributions. Second, we cannot simply choose whether or not to form beliefs on the basis of evidence encountered online. Even if we use social media principally to connect with friends and family members, we will inevitably

encounter information that shapes our beliefs, whether we like it or not (Marin, 2022).

Alternatively, one might think that the problem of fake persons might be resolved by developing our abilities to distinguish between fake and real persons. However, even those possessed of such an ability would struggle to apply it to resolve the challenges raised here. For example, when one is confronted with a dubious tweet with tens of thousands of likes, one cannot realistically examine the Tweet to determine whether its apparent popularity is due only to fake persons. A more feasible alternative to this individualistic suggestion will likely involve technological solutions that make readily available information concerning the extent to which the spread of content online is due to fake persons. Such an alternative holds some promise of helping individuals give evidence its proper weight, thereby avoiding both deception and excessive skepticism.

Acknowledgements I would like to thank two referees for this journal for their exceptionally thoughtful, thorough, and constructive feedback, which helped me to improve the paper considerably.

Author Contribution KRH is the sole author of the content of this manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This is a publication in the context of INTERACT, funded by the Ministry of Culture and Science of North Rhine Westphalia.

Data Availability N/A

Declarations

Ethics Approval and Consent to Participate N/A

Consent for Publication N/A

Competing Interests The author declares no competing interests.

References

Adler, J. (1996). Transmitting knowledge. *Noûs., 31*(1), 99–111.

Alsmadi, I., & O'Brien, M. (2020). How many bots in Russian troll tweets? *Information Processing and Management, 57*(6).

Bastos, M., & Farkas, J. (2019). "Donald Trump is my president!": The Internet Research Agency propaganda machine. *Social Media + Society, 5*(3), 1–13.

Bastos, M., & Mercea, D. (2018). The public accountability of social platforms: Lessons from a study on bots and trolls in the Brexit campaign. *Philosophical Transactions of the Royal Society A, 376*, 1–12.

Bastos, M., & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review, 37*(1), 38–54.

Bernecker, S., Flowerree, A. K., & Grundmann, T. (2021). *The epistemology of fake news*. Oxford University Press.

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

Blake-Turner, C. (2020). Fake news, relevant alternatives, and the degradation of our epistemic environment. *Inquiry.* https://doi.org/10.1080/0020174X.2020.1725623

Broniatowski, D. A., Jamison, A. M., Qi, S. H., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health, 108*, 1378–1384.

Caldarelli, G., De Nicola, R., Del Vigna, F., Petrocchi, M., & Saracco, F. (2020). The role of bot squads in the political propaganda on Twitter. *Communications on Physics, 3*(1), 1–15.

Chen, A. (2015). The agency. The New York Times Magazine. 2/6/2015. https://www.nytimes.com/2015/06/07/magazine/the-agency.html

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of twitter accounts: are you a human bot or cyborg? *IEEE Transactions on Dependable and Secure Computing, 9*(6), 811–824. https://doi.org/10.1109/TDSC.2012.75

Cosentino, G. (2020). *Social media and the post-truth world order: The global dynamics of disinformation*. Springer International Publishing. https://doi.org/10.1007/978-3-030-43005-4

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

Daniel, F., & Millimaggi, A. (2020). On Twitter bots behaving badly: A manual and automated analysis of Python code patterns on GitHub. *Journal of Web Engineering, 18*(8), 801–836.

District Court, U. S. (2018). *United States of America versus internet research agency LLC, Case 1:18-cr-00032-DLFFiled C.F.R. (pp. 1–37). United States District Court for the District of Columbia*.

Eady, G., Paskhalis, T., Zilinsky, J., Bonneau, R., Nagler, J., & Tucker, J. A. (2023). Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nature Communications, 14*(1), 62. https://doi.org/10.1038/s41467-022-35576-9

Ellul, J. (1973). *Propaganda: The formation of men's attitudes*. Vintage Books.

Fallis, D. (2021). The epistemic threat of deepfakes. *Philosophy & Technology., 34*, 623–643. https://doi.org/10.1007/s13347-020-00419-2

Fallis, D., & Mathieson, K. (2019). Fake news is counterfeit news. *Inquiry.* https://doi.org/10.1080/0020174X.2019.1688179

Ferrara, E. (2020). What types of COVID-19 conspiracies are populated by Twitter bots? *First Monday., 25*(6). https://doi.org/10.5210/fm.v25i6.10633

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

Fichman, P., & Sanfilippo, M. R. (2016). *Online trolling and its perpetrators: Under the Cyberbridge*. Rowman & Littlefield.

Fornacciari, P., Mordonini, M., Poggi, A., Sani, L., & Tomaiuolo, M. (2018). A holistic system for troll detection on Twitter. *Computers in Human Behavior, 89*, 258–268.

Gelfert, A. (2018). Fake news: A definition. *Informal Logic, 38*(1), 84–117. https://doi.org/10.22329/il.v38i1.5068

Goldberg, S. (2007). How lucky can you get? *Synthese, 158*, 315–327.

Golovchenko, Y., Buntain, C., Eady, G., Brown, M. A., & Tucker, J. A. (2020). Cross-platform state propaganda: Russian trolls on Twitter and YouTube during the 2016 U.S. Presidential election. *The International Journal of Press/Politics, 25*(3), 357–389.

Graham, P. (2000). Transferring knowledge. *Noûs., 34*, 131–152.

Grundmann, T. (2020). Fake news: The case for a purely consumer-oriented explication. *Inquiry*. https://doi.org/10.1080/0020174X.2020.1813195

Hampton, R. (2019). The Black feminists who saw the Alt-Right threat coming. Slate. https://slate.com/technology/2019/04/black-feminists-alt-right-twitter-gamergate.html

Hannon, M., & de Ridder, J. (2021). The point of political belief. In M. Hannon & J. de Ridder (Eds.), *Routledge Handbook of Political Epistemology* (pp. 156–166). Routledge.
Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research, 6*(2), 215–242.
Hardwig, J. (1985). Epistemic dependence. *Journal of Philosophy, 82*(7), 335–349.
Harman, G. (1973). *Thought*. Princeton University Press.
Harris, K. R. (2021). Video on demand: What deepfakes do and how they harm. *Synthese, 199*(5–6), 13373–13391.
Harris, K. R. (2022). Real fakes: The epistemology of online misinformation. *Philosophy & Technology, 35*(3), 83. https://doi.org/10.1007/s13347-022-00581-9
Jaster, R., & Lanius, D. (2018). What is fake news? *Versus, 2*(127), 207–227.
Jessen, J., & Jørgensen, A. H. (2012). Aggregated trustworthiness: Redefining online credibility through social validation. *First Monday, 17*, 1–2. https://doi.org/10.5210/fm.v17i1.3731
Jones, M. (2019). Propaganda, fake news, and fake trends: The weaponization of Twitter bots in the Gulf crisis. *International Journal of Communication, 13*, 1389–1415.
Kleemans, M., Daalmans, S., Carbaat, I., & Anschütz, D. (2018). Picture perfect: The direct effect of manipulated Instagram photos on body image in adolescent girls. *Media Psychology, 21*(1), 93–110. https://doi.org/10.1080/15213269.2016.1257392
Knustad, M. (2020). Get lost, troll: How accusations of trolling in newspaper comment sections affect the debate. *First Monday., 25*(8). https://doi.org/10.5210/fm.v25i8.10270
Koslowska, H. (2020). Russian trolls and bots are successful because we know they exist. *Quartz*. 30/1/2020. https://qz.com/1792155/russian-trolls-and-bots-are-successful-because-we-know-they-exist/
Krappitz, S. (2012). *Troll culture*. Merz Academy College of Design, Art and Media.
Lackey, J. (2008). *Learning from words: Testimony as a source of knowledge*. Cambridge University Press.
Lackey, J. (2021). Echo chambers, fake news, and social epistemology. In S. Bernecker, A. K. Flowerrese, & T. Grundmann (Eds.), *The epistemology of fake news* (pp. 206–227). Oxford University Press.
Levy, N. (2017). The bad news about fake news. *Social Epistemology Review and Reply Collective, 6*(8), 20–36.
Levy, N. (2022). *Bad beliefs: Why they happen to good people*. Oxford University Press.
Linvell, D. L., & Warren, P. L. (2020). Troll factories: Manufacturing specialized disinformation on Twitter. *Political Communication, 37*(4), 447–467.
Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues. *Communication Research, 49*(2), 171–195. https://doi.org/10.1177/0093650220921321
Marin, L. (2021). Sharing (mis) information on social networking sites. An exploration of the norms for distributing content authored by others. *Ethics and Information Technology, 23*(3), 363–372.
Marin, L. (2022). How to do things with information online. A conceptual framework for evaluating social networking platforms as epistemic environments. *Philosophy & Technology, 35*(3), 77. https://doi.org/10.1007/s13347-022-00569-5
Marsili, N. (2021). Retweeting: Its linguistic and epistemic value. *Synthese, 198*, 10457–10483.
Martini, F., Samula, P., Keller, T. R., & Klinger, U. (2021). Bot, or not? Comparing three methods for detecting social bots in five political discourses. *Big Data & Society*. https://doi.org/10.1177/20539517211033566
McDonald, L. (2021). Please like this paper. *Philosophy, 96*(3), 335–358. https://doi.org/10.1017/S0031819121000152
Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton University Press. https://doi.org/10.1515/9780691198842
Morrison, S. (2021). The rise of the Kremlin troll. *Journal of Media and Information Warfare., 14*(2), 1–14.
O'Sullivan, D. (2017). A notorious Russian Twitter troll came back, and for a week Twitter did nothing. CNN Business, 19/11/2017. https://money.cnn.com/2017/11/17/media/new-jenna-abrams-account-twitter-russia/index.html
Paul, C., & Matthews, M. (2016). The Russian "firehose of falsehood" propaganda model: Why it might work and options to counter it. *RAND Corporation.*, 1–16.
Pomerantsev, P. (2014). Russia and the menace of unreality. The Atlantic. 9/9/2014. https://www.theatlantic.com/international/archive/2014/09/russia-putin-revolutionizing-information-warfare/379880/
Prier, J. (2017). Commanding the trend: Social media as information warfare. *Strategic Studies Quarterly., 11*(4), 50–85.

Rashid, N. (2017). The emergence of the White troll behind a Black face. NPR. https://www.npr.org/sections/codeswitch/2017/03/21/520522240/the-emergence-of-the-white-troll-behind-a-black-face

Rini, R. (2017). Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal, 27*(2), e43–e64.

Rini, R. (2020). Deepfakes and the epistemic backstop. *Philosophers' Imprint., 20*(24), 1–16.

Rini, R. (2021). Weaponized skepticism: An analysis of social media deception as applied political epistemology. In E. Edenberg & M. Hannon (Eds.), *Political epistemology* (pp. 31–48). Oxford University Press.

Ross, R. M., & Levy, N. (2023). Expressive responding in support of Donald Trump: An extended replication of Schaffner and Luks (2018). *Collabra: Psychology, 9*(1), 68054. https://doi.org/10.1525/collabra.68054

Schaffner, B. F., & Luks, S. (2018). Misinformation or expressive responding? *Public Opinion Quarterly, 82*(1), 135–147. https://doi.org/10.1093/poq/nfx042

Schmitt, F. (2017). Social epistemology. In J. Greco & E. Sosa (Eds.), *The Blackwell guide to epistemology* (pp. 354–382). Blackwell Publishing.

Shachaf, P., & Hara, N. (2010). Beyond vandalism: Wikipedia trolls. *Journal of Information Science, 36*(3), 357–370. https://doi.org/10.1177/0165551510365390

Stella, M., Cristoforetti, M., & De Domenico, M. (2019). Influence of augmented humans in online interactions during voting events. *PLoS One, 14*(5), e0214210. https://doi.org/10.1371/journal.pone.0214210

Stukal, D., Sanovich, S., Bonneau, R., & Tucker, J. A. (2022). Why botter: How pro-government bots fight opposition in Russia. *American Political Science Review, 116*(3), 843–857.

Tiku, N. (2022). The Google engineer who thinks the company's AI has come to life. Washington Post. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

Véliz, C. (2023). Chatbots shouldn't use emojis. *Nature, 615*(7952), 375–375. https://doi.org/10.1038/d41586-023-00758-y

Wojcik, S., Messing, S., Smith, A., Rainie, L., & Hitlin, P. (2018). Bots in the Twittersphere. Pew Research Center. 9/4/2018. https://www.pewresearch.org/internet/2018/04/09/bots-in-the-twittersphere