NSP FELLOWS / REVISED PAPER / 30 SEPTEMBER 2026

The Cognitive Siege

How AI and social media change the conditions of information warfare

Samuel Abinsinguza.

Abstract

Social media and generative artificial intelligence can make organized deception cheaper, more adaptable, and easier to scale. Their strategic importance nevertheless depends on a chain of uncertain effects: producing content is different from reaching an audience, changing beliefs, influencing behavior, or weakening institutions. This paper uses "cognitive siege" as an organizing metaphor for sustained, coordinated attempts to manipulate those conditions of public judgment. Drawing on documented influence operations and experimental research, it argues that the most defensible response combines disruption of deceptive infrastructure, accountable platform design, independent research, and public resilience. Democratic governments should address identifiable conduct and measurable risks while protecting lawful disagreement. Neither technological alarmism nor blanket skepticism provides an adequate basis for policy.

The strategic problem

During a six-week period spanning late 2023 and early 2024, Germany's Federal Foreign Office identified more than 50,000 inauthentic accounts on X that produced over 1.8 million automated German-language posts. Its technical report connected the activity to the Doppelganger disinformation campaign. The operation illustrates how a network can manufacture the appearance of many independent voices at a volume that no ordinary reader could investigate individually. The report cautions that view counts include the campaign's own bots and cannot clearly establish genuine audience reach. Output at this scale is evidence of an organized operation, but not of how many people believed it or changed their behavior.[1]

That distinction is central to understanding contemporary information warfare. State-linked operators and other organized actors can exploit digital platforms to impersonate trusted sources, conceal sponsorship, and insert repeated narratives into public debate. Recommendation systems can then affect which messages users encounter. Generative AI can support content production and adaptation. These capabilities create serious risks, but an account of their strategic consequences must explain the steps between an operator's activity and an audience's response.

I use "cognitive siege" here as an organizing metaphor for sustained, coordinated efforts to manipulate the information through which people interpret events and make decisions. The phrase is not presented as an established doctrine or a newly validated theory. It emphasizes cumulative pressure on public judgment: repeated exposure, misleading cues of credibility, and attempts to

exploit existing uncertainty. A campaign need not persuade everyone to matter. It might seek to mobilize a receptive minority, confuse attribution, divert institutional attention, or make a policy harder to sustain. Those are possible objectives, however, and each requires evidence of its own.

For this paper, an influence operation is an organized effort to affect an audience's perceptions or behavior in pursuit of an objective. Information warfare refers more narrowly to the strategic use of such efforts in political or security competition. These working definitions do not make every controversial message an act of war. Ordinary political advocacy, independent journalism, satire, and organic protest can all be provocative or misleading without belonging to a coordinated operation. Attribution should rest on evidence about actors, coordination, sponsorship, deceptive identities, and operational objectives, rather than on whether an analyst dislikes the narrative.

The core argument is that AI changes the cost and organization of influence more clearly than it changes the underlying psychology of persuasion. Policy should therefore concentrate on the mechanisms that are observable and contestable: deceptive infrastructure, exploitative data practices, distribution systems, and weaknesses in independent scrutiny. Claims of lasting social or geopolitical effects deserve a higher evidentiary threshold than claims that a campaign produced a large number of posts. This paper is an analytical synthesis of selected documented operations and experiments, with sources checked in September 2026. It is not a systematic review or a new empirical study.

Old ambitions and changing capabilities

Propaganda long predates social media. Efforts to influence public opinion through repetition, selective evidence, fabricated sources, and concealed sponsorship also long predate generative AI. A CIA assessment from 1986 described Soviet disinformation falsely linking the United States to the creation and spread of AIDS. The historical example matters because it shows that source laundering and attacks on institutional credibility did not require digital recommendation systems.[2]

The same continuity is visible in the Internet Research Agency's activity around the 2016 United States election. The Senate Intelligence Committee documented a Russian influence effort that used false American personas, social-media accounts, and divisive themes. This occurred before the public deployment of contemporary general-purpose language models. It would therefore be inaccurate to describe fake accounts, audience segmentation, or long-term efforts to alter beliefs as inventions of the current AI era.[3]

Christopher Paul and Miriam Matthews' analysis of Russia's "firehose of falsehood" model offers a useful account of high-volume, multichannel, repetitive propaganda. Their argument helps explain why operators may value saturation and the appearance of numerous sources even when individual messages are inconsistent. It should not be treated as proof that every large campaign succeeds. A description of an influence strategy and evidence that it achieved a political outcome are different kinds of claim.[4]

What, then, changes with generative AI? The first change is in production. Language models can draft, translate, and revise material across formats, reducing some of the language and writing work

involved in maintaining a campaign. In its 2024 investigation of Storm-2035, OpenAI observed accounts using ChatGPT to generate articles and social-media comments for a covert Iranian influence operation. This provides concrete evidence of AI entering an existing operational workflow.[5]

The second change is the possibility of adaptation. An operator can generate variants for different audiences or use conversational systems to respond to a reader rather than distribute only a fixed message. Laboratory studies demonstrate that AI-generated political messages and dialogues can influence reported attitudes under some conditions. They do not establish that a campaign can reliably infer an individual's vulnerabilities, reach that person in daily life, and produce a durable behavioral change. Those additional steps should be investigated rather than assumed.[6]

The third change concerns the relationship between production and distribution. More plentiful content gives operators additional material to test, select, and place into existing channels. Platform feedback may help them identify what attracts attention. Yet distribution remains a constraint: operators still need credible accounts, access to relevant audiences, persistence through enforcement, and messages that connect with local concerns. AI can reduce one bottleneck while leaving others intact. The important comparison is with the cost and performance of a plausible human-led campaign, not with an imaginary world without propaganda.

Digital media can also strengthen resistance. A reader who once depended on a small number of outlets may now compare reporting, inspect original records, encounter corrections, and challenge an official narrative. Greater media plurality can reduce dependence on a single source. The risk emerges when apparent plurality conceals common control, or when the effort required to establish source independence exceeds the attention a reader can reasonably give. Neither greater access nor greater information volume has a uniformly beneficial or harmful effect.

A study linking surveys to the Twitter feeds of 1,496 United States users during the 2016 election gives this counterargument empirical weight. Potential exposure to Internet Research Agency content was highly concentrated and dwarfed by content from domestic news outlets and politicians. The researchers found no meaningful relationship between that exposure and measured changes in attitudes or voting behavior. The observational design and reconstructed potential exposure limit causal conclusions; the study also cannot settle the effects of other platforms or interference methods.[7] It nonetheless shows why an operation must be assessed within the wider information people receive.

From content production to strategic consequences

An adequate analysis separates five questions: what was produced, who encountered it, what they came to believe, how they acted, and whether those actions changed an institutional or strategic outcome. This is an analytical distinction, not a claim to a new causal theory. Its value is practical: it prevents evidence at one stage from silently doing the work of evidence at every later stage.

Production is often the easiest stage to document. Investigators can count accounts, domains, posts, or generated images and reconstruct their coordination. These measures establish activity and

operational capacity. They do not establish audience size. A thousand accounts can circulate messages among themselves; a single well-connected intermediary can sometimes bring a message to a much larger audience. Counts also need context: how long the activity lasted, whether accounts were automated, and how much authentic traffic existed in the same environment.

Exposure concerns who actually saw a message, how often, and alongside what else. Views and impressions are more informative than production counts, but they can include repeat encounters, inattentive viewing, or engagement by opponents. Recommendation systems, paid promotion, subscriptions, interpersonal sharing, and news coverage can each affect distribution. Analysts should avoid attributing every encounter to "the algorithm" when the route of exposure has not been established.

Persuasion concerns changes in belief, interpretation, issue importance, or attitude. Psychological research offers plausible mechanisms, including greater perceived accuracy after repetition. For example, Pennycook, Cannon, and Rand found that prior exposure increased the perceived accuracy of false news headlines in their experiments. This supports a bounded claim about familiarity and judgments of accuracy; it does not show that any repeated claim will overcome prior commitments or determine a vote.[8] Similarly, repeated messages from apparently independent outlets may create an impression of corroboration. Genuine independence, source recognition, existing knowledge, and attention all affect how that cue is interpreted.

Experimental platform research complicates simple claims about polarization. In a 2020-election study, researchers reduced exposure to like-minded sources on Facebook by about one-third among 23,377 participants. They found no measurable effects on the eight preregistered attitudinal outcomes, including affective polarization and ideological extremity. This does not establish that information environments never matter. It shows that a substantial change in one type of exposure did not produce the anticipated attitudinal changes in that setting and observation period.[9]

A different experiment on X, conducted in 2023 and published in 2026, distinguished users by their initial feed preference. Among initial chronological-feed users, assignment to the algorithmic feed shifted some current-news and issue attitudes. Among initial algorithmic-feed users, switching it off produced no detectable corresponding effects. Neither switch significantly changed affective polarization or party identification. Because the initial preference groups were not randomized, differences between them may contribute to the asymmetry. The Facebook and X studies also differ in platform, timing, and intervention. Together, they caution against treating all beliefs as equally responsive or assuming that a feed change must either transform political identity or have no effect at all.[10]

Research on AI persuasion needs the same discipline. A 2025 randomized study of short online debates found that personalized GPT-4 was more persuasive than a human baseline under its experimental conditions. The paper's 2026 correction clarifies that the difference between personalized and non-personalized GPT-4 did not reach statistical significance. A separate large experiment on AI-generated political messages also found no significant persuasive advantage from the tested microtargeting approach.[11] These results support taking AI persuasion seriously while

resisting a stronger claim that personalization has already been shown to confer a reliable advantage across contexts.

Behavior is a further step. Agreement with a statement on a survey is different from donating, attending a protest, refusing a vaccine, or changing a vote. People may express views they do not act on, face practical constraints, or already hold a position close to the message. Studies of behavioral effects should specify the behavior, its timing, and a credible comparison group. Engagement should not automatically stand in for endorsement: sharing can express outrage, mockery, or an attempt to correct a falsehood.

Finally, institutional and strategic outcomes include public confidence, the implementation of policy, election results, and the endurance of alliances. These emerge from many causes. Economic conditions, political leadership, genuine grievances, armed conflict, and institutional performance can influence the same outcomes an operation seeks to affect. A campaign's stated ambition is evidence about intent. Establishing its contribution to an outcome requires stronger causal evidence. Uncertainty at this stage is a reason to improve research and proportionate defenses, rather than a license to declare either decisive success or harmlessness.

What documented operations can establish

Doppelganger illustrates deception in source identity and distribution. In 2024, the United States Department of Justice announced the seizure of 32 internet domains associated with the operation. The supporting affidavit alleged that actors working under Russian government direction used domains that imitated legitimate news outlets and sought to advance Russian policy interests. That evidence supports claims about impersonation, operational organization, and alleged sponsorship. It does not by itself demonstrate that the operation changed an election result or persuaded its intended audience.[12]

The German technical findings complement that account by documenting extensive automated posting. The analytical lesson is not simply that propaganda has become larger. It is that source imitation and coordinated repetition can obscure whether multiple apparent endorsements are genuinely independent. Responses such as disrupting impersonating domains and tracing coordinated account networks address that deception more directly than an instruction to suppress an entire political viewpoint.

Storm-2035 provides a useful test of the assumption that AI-generated output automatically translates into influence. OpenAI reported that the operation used ChatGPT for long-form articles and shorter English- and Spanish-language comments. It also reported that the identified activity did not achieve meaningful audience engagement: most of the observed social-media posts received few or no likes, shares, or comments. The case demonstrates adoption of AI for production, without evidence of meaningful audience engagement in the activity the provider observed.[13]

Vendor investigations are valuable because providers can see activity unavailable to outside researchers. They are also partial views, bounded by the services, accounts, and methods an investigator can observe. Their findings should be attributed to the reporting organization and,

where possible, compared with independent evidence. Microsoft's 2024 reporting on China-linked influence activity likewise described AI-enhanced content while acknowledging little evidence that the observed efforts had successfully swayed opinion.[14] Such qualifications belong in an assessment of the threat, not in a footnote to an otherwise conclusive narrative.

Cambridge Analytica illustrates a related but distinct problem. The Federal Trade Commission's action documented deceptive practices connected with collecting Facebook information for voter profiling and targeting. The episode is strong evidence of a privacy and accountability failure. It is a weaker basis for claims that psychographic targeting reliably changed political behavior: the possession of extensive personal data and a company's marketing claims do not independently establish persuasive effectiveness.[15] Privacy safeguards can be justified by unlawful or deceptive data practices without requiring proof that a particular election was altered.

These cases also show why origin and mechanism should be stated precisely. Evidence about one Russian-linked operation cannot establish the methods or success of a Chinese or Iranian one. Domestic actors can use similar techniques, and authentic participants may amplify material without knowing its origin. The relevant unit of analysis is the documented operation and its observable conduct, not a generalized category of foreign manipulation.

Governing manipulation without governing belief

A response that empowers authorities to decide which political claims may circulate could damage the democratic judgment it seeks to protect. Definitions of disinformation are contested, evidence develops, and governments have incentives to classify inconvenient criticism as manipulation. Platforms can also make consequential mistakes. Free-expression concerns are therefore a central design constraint, rather than an obstacle to be mentioned after recommending stronger enforcement.

The United States and the European Union illustrate different legal settings. In Moody v. NetChoice, the United States Supreme Court explained that selecting and arranging third-party content can involve protected editorial judgment. The Court returned the facial challenges for further analysis; its decision is not a categorical ruling that every platform function or transparency requirement is constitutionally immune from regulation.[16] It does, however, make a simple government mandate to carry or suppress specified political messages a poor foundation for a general policy proposal. Informal cooperation also needs safeguards: officials must not use coercive threats against intermediaries to suppress disfavored speech, a principle the Court addressed in NRA v. Vullo.[17] Government–platform communications should distinguish voluntary notices from binding orders, record their legal basis, and permit independent review.

The EU Digital Services Act offers a different, risk-based approach. It includes obligations for very large platforms and search engines concerning systemic-risk assessment, mitigation, independent audits, and researcher access. Its framework addresses risks to civic discourse and electoral processes alongside other protected interests. The Act does not make all false or disputed political speech illegal. Its mitigation provisions also require attention to fundamental rights, including

freedom of expression.[18] Transplanting that framework into another jurisdiction would require local constitutional and institutional analysis.

Within those constraints, four priorities follow from the mechanisms examined here.

First, platforms and competent authorities should act against demonstrable deception in identity and infrastructure. Impersonation, coordinated inauthentic behavior, concealed sponsorship, and unauthorized data acquisition provide more concrete grounds for investigation than the popularity or ideology of a claim. Platforms should explain the relevant rule, preserve evidence, distinguish coordinated operations from legitimate collective action, and provide meaningful appeal. Law-enforcement action requires the applicable legal basis and oversight. Anonymity should not be treated as evidence of deception by itself; journalists, dissidents, and vulnerable participants may have legitimate reasons to protect their identities.

Second, accountability should focus on how distribution systems operate. Qualified independent researchers need ways to examine exposure, recommendation, advertising, and enforcement without receiving unrestricted personal information. Secure access arrangements, data minimization, clear research eligibility, and enforceable confidentiality obligations can reduce the tension between scrutiny and privacy. The DSA's researcher-access provisions provide a concrete institutional model for this work.[19] Publishing source code alone would not show what users actually encountered, and it could expose security-sensitive details without answering the central causal questions. Audit access should be designed around those questions.

Third, users need intelligible choices and accurate context. Platforms should make recommendation controls understandable, explain significant targeting criteria, and enable users to see who paid for political advertising. Labels and fact-checks should identify a specific verifiable claim, explain the evidence, and distinguish factual correction from opinion or prediction. Coverage will inevitably be incomplete, especially across languages and rapidly changing events. Appeals, visible corrections, and independent evaluation are necessary because inaccurate labels can themselves undermine confidence. Legal duties should specify procedures and accountability where justified, rather than presume that a mandatory fact-checking quota can produce an authoritative inventory of truth.

Fourth, interventions should be proportionate to the harm and the evidence. A fabricated emergency instruction or impersonated election authority may warrant a different response from a disputed assessment of foreign policy. Restrictions on an entire service can impose substantial costs on lawful speech, association, and access to information. Such measures require a separate, specific justification and consideration of narrower alternatives; their existence is not evidence that broad restrictions are an effective remedy for misinformation. Authorities should publish the legal basis and provide review mechanisms appropriate to the action.

This approach does not eliminate difficult judgments. Rules against inauthentic coordination can be evaded or applied too broadly; disclosure can become surveillance; researcher access can create privacy risks; and moderation appeals can be too slow to remedy a time-sensitive mistake. Those tradeoffs should be designed into implementation. The test is whether a measure reduces an

identified risk at a defensible cost to rights, privacy, and legitimate participation, with evidence that outsiders can examine.

Building resilience and testing results

Public resilience requires more than asking individuals to be skeptical of everything. Indiscriminate distrust can make reliable reporting and fabricated material seem equally suspect. The practical objective is calibrated judgment: checking a claim against independent evidence, noticing impersonation, understanding uncertainty, and knowing when a source has corrected an error.

There is experimental support for modest interventions. Research on accuracy prompts suggests that briefly drawing attention to accuracy can improve the quality of information people choose to share. A media-literacy experiment improved discernment between mainstream and false news in United States and educated online Indian samples, but not in a rural Indian sample. The intervention also lowered perceived accuracy of mainstream reporting and did not measurably change news-site consumption.[20] Neither approach substitutes for platform responsibility, and neither should be presented as a universal or permanent defense. Implementation should test persistence, language coverage, accessibility, and whether gains survive outside an experiment.

Education can build on those findings through concrete practices. Students and adult learners can compare a claim with its original source, distinguish corroboration from repeated copying, and trace the provenance of an image or quotation. Public institutions should make reliable primary information easy to locate, explain changes in their assessments, and correct mistakes promptly. Trust is not sustained solely by teaching citizens to identify misleading messages; institutions must also behave in ways that merit scrutiny and confidence.

Interdisciplinary collaboration becomes useful when it produces testable tools and publicly assessable results. DARPA's Social Media in Strategic Communication program, for example, supported research on detecting and analyzing online influence. A documented output was the 2015 DARPA Twitter Bot Challenge, in which teams combined computational analysis with human review to identify 39 known influence bots in a replayed Twitter data set. The three top teams identified all 39 organizer-known bots; the controlled benchmark supplied feedback on submitted guesses, making it a concrete, testable detection exercise.[21] It did not establish effectiveness against later adversaries or demonstrate reduced public susceptibility to manipulation.

Evaluation should follow the same distinctions used to assess the threat. Detection tools can be tested for false positives, false negatives, and performance across languages. Enforcement can be assessed for how long deceptive networks remain active and whether they simply migrate. Distribution interventions can be examined for actual changes in exposure. Persuasion and resilience measures require tests of relevant beliefs or behaviors, including persistence. Institutional consequences need an appropriate comparison and a realistic time horizon.

Platforms, researchers, and public bodies should also report failures. A campaign that generates thousands of posts but attracts little attention, or a well-designed intervention that changes exposure without changing attitudes, is not an uninformative result. Such findings locate

constraints, help prevent exaggerated threat assessments, and direct resources toward mechanisms that warrant closer study. Success should not be defined by the number of accounts removed, the volume of content labeled, or an increase in generalized public trust alone.

Conclusion

The cognitive siege is a useful metaphor only if it sharpens analysis rather than substitutes for it. Social media can conceal common control behind apparently independent sources; recommendation systems can alter exposure; generative AI can expand the supply and adaptability of persuasive content. None of those facts establishes an irresistible ability to control public opinion.

The strategic concern is the combination of organized deception, opaque distribution, and repeated pressure on public judgment. Its consequences vary with the audience, the issue, the surrounding information environment, and the institutions asked to respond. The strongest case for action therefore combines historical perspective with causal restraint.

Democratic resilience requires the capacity to investigate manipulation without treating disagreement as a security threat. Disrupting deceptive infrastructure, protecting personal information, enabling independent scrutiny, and testing practical interventions are defensible starting points. The aim is to make public judgment harder to covertly manipulate while preserving the freedom to challenge both private platforms and the state.

Notes and sources

1. German Federal Foreign Office, Technical Report on the Doppelganger Disinformation Campaign (2024), pp. 4, 12–13, 28 (German).

2. Central Intelligence Agency, 'USSR: Increasing Disinformation About AIDS' (1986), declassified assessment, Ronald Reagan Presidential Library archive, PDF p. 37 (original p. 12); year follows archive metadata and internal chronology.

3. U.S. Senate Select Committee on Intelligence, Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, vol. 2, Russia's Use of Social Media with Additional Views (2019), pp. 3–5.

4. Christopher Paul and Miriam Matthews, The Russian 'Firehose of Falsehood' Propaganda Model: Why It Might Work and Options to Counter It, RAND Perspective PE-198-OSD (2016).

5. OpenAI, 'Disrupting a Covert Iranian Influence Operation' (2024), provider investigation of Storm-2035, including observed engagement limits.

6. Francesco Salvi, Manoel Horta Ribeiro, Riccardo Gallotti, and Robert West, 'On the Conversational Persuasiveness of GPT-4,' Nature Human Behaviour 9 (2025): 1645–1653. See the authors' 2026 correction of the personalization comparison. Kobi Hackenburg and Helen Margetts, 'Evaluating the Persuasive Influence of Political Microtargeting with Large Language Models,' PNAS 121, no. 24 (2024): e2403116121.

7. Gregory Eady et al., '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 (2023): 62.

8. Gordon Pennycook, Tyrone D. Cannon, and David G. Rand, 'Prior Exposure Increases Perceived Accuracy of Fake News,' Journal of Experimental Psychology: General 147, no. 12 (2018): 1865–1880.

9. Brendan Nyhan et al., 'Like-minded Sources on Facebook Are Prevalent but Not Polarizing,' Nature 620 (2023): 137–144.

10. Germain Gauthier, Roland Hodler, Philine Widmer, and Ekaterina Zhuravskaya, 'The Political Effects of X's Feed Algorithm,' Nature 652 (2026): 416–423, especially the experimental design and limitations of comparing initial feed-preference groups.

11. Francesco Salvi, Manoel Horta Ribeiro, Riccardo Gallotti, and Robert West, 'On the Conversational Persuasiveness of GPT-4,' Nature Human Behaviour 9 (2025): 1645–1653. See the authors' 2026 correction of the personalization comparison. Kobi Hackenburg and Helen Margetts, 'Evaluating the Persuasive Influence of Political Microtargeting with Large Language Models,' PNAS 121, no. 24 (2024): e2403116121.

12. U.S. Department of Justice, 'Justice Department Disrupts Covert Russian Government-Sponsored Foreign Malign Influence Operation' (2024). See the supporting Doppelganger affidavit, especially pp. 1–3 and 15, for allegations concerning direction and impersonation.

13. OpenAI, 'Disrupting a Covert Iranian Influence Operation' (2024), provider investigation of Storm-2035, including observed engagement limits.

14. Microsoft, 'China Tests US Voter Fault Lines and Ramps AI Content to Boost Its Geopolitical Interests' (2024), Microsoft Threat Analysis Center findings and stated limits on evidence of persuasion.

15. Federal Trade Commission, 'FTC Issues Opinion and Order Against Cambridge Analytica for Deceiving Consumers About the Collection of Facebook Data, Compliance with EU-U.S. Privacy Shield' (2019). See the Commission's final opinion, In re Cambridge Analytica, LLC, Docket No. 9389, pp. 4–8.

16. Moody v. NetChoice, LLC, 603 U.S. 707 (2024), majority opinion, especially slip-opinion pp. 2–4 and 17–29 (vacating and remanding the facial challenges).

17. National Rifle Association of America v. Vullo, 602 U.S. 175 (2024), holding that the alleged coercive conduct plausibly stated a First Amendment claim at the motion-to-dismiss stage.

18. European Union, Regulation (EU) 2022/2065 (Digital Services Act), especially Articles 17, 20–21, 27, 34–35, 37–38, and 40. European Commission, 'FAQs on DSA Data Access for Researchers' (2025), on public-data and vetted-researcher access, safeguards, and eligibility.

19. European Union, Regulation (EU) 2022/2065 (Digital Services Act), especially Articles 17, 20–21, 27, 34–35, 37–38, and 40. European Commission, 'FAQs on DSA Data Access for Researchers' (2025), on public-data and vetted-researcher access, safeguards, and eligibility.

20. Gordon Pennycook et al., 'Shifting Attention to Accuracy Can Reduce Misinformation Online,' Nature 592 (2021): 590–595. Andrew M. Guess et al., 'A Digital Media Literacy Intervention Increases Discernment Between Mainstream and False News in the United States and India,' PNAS 117, no. 27 (2020): 15536–15545.

21. V. S. Subrahmanian et al., 'The DARPA Twitter Bot Challenge,' Computer 49, no. 6 (2016): 38–46; author preprint, pp. 2–4, on the 2015 controlled benchmark and team results.