THE GLOBAL RACE FOR TECHNOLOGICAL SUPERIORITY

Discover the Security Implications

edited by **Fabio Rugge**
introduction by **John R. Allen** and **Giampiero Massolo**

ISPI
BROOKINGS

THE GLOBAL RACE FOR TECHNOLOGICAL SUPERIORITY

edited by Fabio Rugge

ISPI BROOKINGS

Via Alamanni, 11 – 20141 Milano – Italy
www.ledizioni.it
info@ledizioni.it

The Global Race for Technological Superiority
Edited by Fabio Rugge
First edition: November 2019

This report is published with the support of the Italian Ministry of Foreign Affairs and International Cooperation (in accordance with Article 23- bis of the Decree of the President of the Italian Republic 18/1967), within the framework of the activities of the Centre on Cybersecurity jointly promoted by ISPI and Leonardo. The opinions expressed are those of the authors.

Print ISBN 9788855261432
ePub ISBN 9788855261449
Pdf ISBN 9788855261456
DOI 10.14672/55261432

ISPI. Via Clerici, 5
20121, Milan
www.ispionline.it

Catalogue and reprints information: www.ledizioni.it
Cover image: Fulvio ranieri mariani, turbine aereo, 1938

BROOKINGS

The Brookings Institution is a nonprofit organization devoted to independent research and policy solutions. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public.

Table of Contents

5. Artificial Intelligence, Geopolitics, and Information Integrity

John Villasenor

Much has been written, and rightly so, about the potential that artificial intelligence (AI) can be used to create and promote misinformation. But there is a less well-recognized but equally important application for AI in helping to *detect* misinformation and limit its spread. This dual role will be particularly important in geopolitics, which is closely tied to how governments shape and react to public opinion both within and beyond their borders. And it is important for another reason as well: While nation-state interest in information is certainly not new, the incorporation of AI into the information ecosystem is set to accelerate as machine learning and related technologies experience continued advances.

The present article explores the intersection of AI and information integrity in the specific context of geopolitics. Before addressing that topic further, it is important to underscore that the geopolitical implications of AI go far beyond information. AI will reshape defense, manufacturing, trade, and many other geopolitically-relevant sectors. But information is unique because information flows determine what people know about their own country and the events within it, as well as what they know about events occurring on a global scale. And information flows are also critical inputs to government decisions

regarding defense, national security, and the promotion of economic growth. Thus, a full accounting of how AI will influence geopolitics of necessity requires engaging with its application in the information ecosystem.

This chapter begins with an exploration of some of the key factors that will shape the use of AI in future digital information technologies. It then considers how AI can be applied to both the creation and detection of misinformation. The final section addresses how AI will impact efforts by nation-states to promote – or impede – information integrity.

AI and the Information Ecosystem: Some Key Factors

Advancing AI Technologies

A combination of factors will determine how AI will impact the information ecosystem over the next decade. First, there is the technology itself. Spurred by extraordinary levels of both private and public investment, AI is advancing at far greater rates than in the past. According to CB Insights, venture capital investment in the United States in AI startups grew from $4.1 billion in 2016 to $5.4 billion in 2017 to $9.3 billion in 2018[1]. The US government has also been ramping up its support for AI research. For example, in fall 2018 the US Department of Defense's Defense Advanced Research Projects Agency (DARPA) announced a "$2 billion campaign to develop next wave of AI technologies"[2].

In China, which views AI as a central focus of its goal of becoming a technological superpower, the government has launched a wide array of multi-billion-dollar AI investment

[1] CB Insights, "VCs Nearly Doubled Their Investment in This Tech Last Year", 20 February 2019.

[2] Defense Advanced Research Project Agency, "DARPA Announces $2 Billion Campaign to Develop Next Wave of A.I. Technologies", 7 September 2018.

initiatives[3]. Israel is another key player in the global AI landscape. In 2018, "AI-related companies accounted for 17% of the total number of 6,673 active Israeli tech companies in Israel tracked by Start-Up Nation Finder" and "32% of all funding rounds and 37% of the total capital raised went to AI-related companies"[4]. And in Europe, the European Commission has announced a plan aimed at spurring "more than €20 billion per year from public and private investments" in AI over the 2020s[5].

An additional aspect of the landscape not captured by the statistics above is the enormous internal AI research and development investment being made by large companies such as Amazon, IBM, Google, and Microsoft. Collectively, the capital flowing from governments, venture investors, and corporations will spur extraordinary AI advances, greatly broadening the capacity to analyze and make effective use of data. Relatedly, continued investment will make AI better at learning, opening the door to increasingly sophisticated algorithms that combine human ingenuity with computer-driven insights.

The Growing Role of AI in the Digital Ecosystem

A second factor that will elevate the role of AI is the degree to which it will be increasingly intertwined with broader digital information ecosystem. Many of the most important information technology changes of the last quarter of a century – including the growth of the internet, advances in digital storage and computation capacity, and the introduction and mass adoption of smartphones and social media – have occurred largely (though not completely) without AI. By contrast, the future evolution of the digital information landscape will be driven in significant part by AI.

---

[3] T.H. Davenport, "China Is Executing its for AI while is still wrestling to create one", *Market Watch*, 27 February 2019.

[4] A. Mizroch, "In Israel, A Stand Out Year for Artificial Intelligence Technologies", *Forbes*, 11 March 2019.

[5] AI Europe Hub, "European Union to Invest 20 Billion Euros in AI".

Over about the last five years, we have been experiencing the first stages of this transition, and AI is now used a wide range of commercial products and services. There is an understandable temptation to predict the future by extrapolating the past, and therefore to conclude that the next 5 or 10 years see the introduction of even more AI into the commercial ecosystem to enhance consumer services in areas such as transportation, online purchasing, and media delivery. But while that prediction is no doubt accurate, it almost certainly fails to anticipate the more profound AI-induced changes that are much harder to foresee in advance.

By analogy, consider the internet in the late 1990s. At that time, it would have been relatively easy to predict dramatic growth in both the number and diversity of web sites over the subsequent 10 years. But it would have been much harder to envision the growth and impact of social media—which we now know spurred far more significant changes than did growth in the number of websites. In the same way, it is easy today to conclude that AI will play an increasingly large role in the digital information landscape over the next decade, but far harder to anticipate its use in ways that lack clear historical antecedents.

Information Gatekeepers

Information gatekeepers, including but not limited to social media companies, constitute a third factor influencing how AI will shape the information ecosystem. For large-scale social media companies, as well as other companies (such as online retailers and providers of internet and mobile phone services) that engage with millions of individual users, the question is not whether to incorporate AI, but rather how it should be most effectively used to further goals such as offering highly customized content to consumers and detecting fraud. As AI continues to advance, companies seeking to take advantage of the cost efficiencies it enables have incentives to deploy it more extensively in their systems. Companies will make highly

consequential policy choices regarding their development and rollout of AI solutions, addressing questions such as the extent to which they should curate and/or filter content, the standards they will apply in relation to testing and monitoring algorithms to detect problems such as bias, and the level of human oversight to provide in relation algorithmic decisions and algorithmic evolution.

In authoritarian countries, an additional information gatekeeper is the government itself. All authoritarian governments will seek to use AI to monitor online traffic and detect digital content deemed problematic. But there will be variations both across and within authoritarian countries in the nature of the tools employed and the extent to which they are used to actively control (as opposed to monitor) discourse.

AI and Information Integrity

"Information integrity" as used herein is intended to describe the extent to which information is accurate, non-deceptive, and properly attributed. While accuracy is clearly a baseline requirement to achieve information integrity, accuracy alone will not always be sufficient. For information to have integrity it also has to be contextualized in a manner that avoids deception. To take a simple example, consider a politician who accompanies a family member who has struggled with drug addiction on a visit to a drug rehabilitation clinic. Suppose that the politician is photographed when leaving the clinic, and that those photographs are then distributed on social media. The photographs are accurate in the sense of depicting an event that actually occurred, but they are deceptive because, when distributed without context, they could imply that the politician is personally struggling with drug addiction.

Attribution is also important. A social media posting purporting to come from a voter and containing accurate, properly contextualized content still lacks integrity if in fact it was posted by a foreign government aiming to influence an election. Thus,

challenges to assessing the integrity of information include not only evaluating truth or falsity, but also identifying the extent to which decontextualization may lead to misinterpretation, as well as understanding whether the purported source is the same as the actual source.

Much of the recent public dialog regarding the role of AI in information integrity has focused on potential negative impacts. Deepfakes, which are videos produced with the aid of deep learning techniques that portray people doing or saying things that they never did or said, have been correctly identified as a major potential concern[6]. A well-constructed deepfake targeting a politician, if released onto the internet at the right time and manner, could potentially swing a close election.

AI can also be used to undermine information integrity in other ways. Consider "bots", which describe accounts on Twitter and other social media platforms that masquerade as humans but are actually software (though as of yet, not generally *AI-enabled* software). While precise statistics on the percentage of Twitter accounts that are bots are hard to come by (in part due to fluctuations over time as different bot detection techniques are developed and deployed, and as bot creators then react by updating their methods), it is clear that the number is very high.

Bots are known to play an important role in amplifying online misinformation. A November 2018 paper published in *Nature Communications* reported on a study of "14 million messages spreading 400 thousand articles on Twitter during ten months in 2016 and 2017"[7]. The authors found "evidence that social bots played a disproportionate role in spreading articles from low-credibility sources. Bots amplify such content in the early spreading moments, before an article goes viral. They also

[6] It is important to note that deepfakes are not inherently bad. Deepfakes have plenty of innocuous uses as well, including in areas such as education and entertainment.

[7] C. Shao et al., "The spread of low-credibility content by social bots", *Nature*, vol. 9, 2018.

target users with many followers through replies and mentions. Humans are vulnerable to this manipulation, resharing content posted by bots"[8]. As noted above, in the past, most bots have not been AI-enabled. Inevitably, this will change. Well-designed AI-powered bots could do a very effective job of impersonating humans, making them much harder to detect and more effective at disseminating misinformation.

As concerning as the above examples are, it is also important to consider the other side of the ledger. Just as AI can be used to promote misinformation, it can also be used to combat it. Deepfake detection is one example. There is a very active community of researchers working to develop methods, including approaches based on AI, to automatically identify manipulated videos. Examples include the use of deep learning to identify artifacts introduced by face-swapping software[9] and the use of neural networks to identify frame-to-frame inconsistencies in deepfake videos[10]. As a February 2019 article in *IEEE Spectrum* noted, "the AI Foundation raised $10 million to build a tool that uses both human moderators and machine learning to identify deceptive malicious content such as deepfakes"[11]. The same article also described efforts by a Netherlands-based technology startup to use adversarial machine learning "as a primary tool for detecting deepfakes"[12].

AI can also be used to detect activity by bots. Bots that do not rely on AI often act in recognizable ways that can easily be detected. The authors of the *Nature Communications* article noted above observed that when low-credibility content goes viral, it exhibits "distinctive patterns"[13]. The authors explained that

---

[8] Ibid.

[9] Yuezun Li and Siwei Liu, *Exposing DeepFake Videos by Detecting Face Warping Artifacts*, Working Paper, 22 May 2019.

[10] D. Guera and E.J. Delp, *Deefake Video Detection Using Recurrent Neural Networks*, Working Paper.

[11] J. Hsu, *Can AI Detect DeepFakes to Help Ensure Integrity of U.S. 2020 Elections*, IEEE, 28 February 2019.

[12] Ibid.

[13] C. Shao et al. (2018).

> most articles by low-credibility sources spread through original tweets and retweets, while few are shared in replies; this is different from articles by fact-checking sources, which are shared mainly via retweets but also replies. In other words, the spreading patterns of low-credibility content are less "conversational". Second, the more a story was tweeted, the more the tweets were concentrated in the hands of few accounts, who act as "super-spreaders"[14].

By contrast, in the future when many bots become AI-enabled, they will be more capable of emulating organic, non-coordinated viral behavior, in part by creating larger networks to spread tweets and in part by relying more on including misinformation in "replies" that might appear to have been written by a real person. The most effective way to identify and block AI-enabled bots will be to use AI in the detection algorithms. Such algorithms could monitor the evolving behavior of a bot network, and in response evolve their own templates for identifying likely non-human social media activity.

The examples of deepfakes and bots illustrate that while misinformation poses major challenges, the same powerful AI techniques that can be employed to produce false or deceptive content can also be applied to its detection and mitigation. A challenge is that the asymmetries involved give misinformation creators an inherent set of advantages. They can continually enhance their algorithms to stay one step ahead of the latest detection techniques. And, to have impact, misinformation creators only have to succeed some of the time. Even if only a low percentage of malicious content evades detection, that can still be enough to cause significant harms.

Governments and the Information Ecosystem

As the above discussion makes clear, over the next decade AI will experience dramatic advances and take on an increasing role in the broader digital information ecosystem. At the same

[14] Ibid.

time, AI-based techniques for generating misinformation will become more sophisticated, as will techniques for detecting and impeding its spread.

This will impact geopolitics in multiple important ways. In authoritarian countries, governments have always sought to exert high levels of control over information, both through propagation of state-approved content and censorship of content deemed inconsistent with the government objectives. AI offers a powerful tool for achieving these ends. To take one example, AI can make it easy for an authoritarian country to perform highly detailed inspection and censorship of social media postings. Postings can be examined not only individually, but also in the aggregate for an individual or group of individuals to identify broader trends that might be of interest to the government. Authoritarian governments will make use of these capabilities to further geopolitical (and other) goals.

Inevitably, some governments will also seek to use online misinformation to alter elections in other countries. The well-documented foreign manipulation of US social media to attempt to influence the 2016 US presidential election is, unfortunately, only a foreshadowing of what is likely to occur in future high-stakes elections. AI-powered misinformation aimed at swaying voter perceptions can be very effective. Combating it will be challenging in part because of the high degree of coordination that would be needed among multiple private and public sector entities to identify and mitigate foreign government misinformation. Yet another complicating factor is that some forms of manipulation can be subtle and therefore not easily detectable. For instance, a foreign government might use AI to create social media accounts in the target country and cause those accounts to engage in much more humanlike behavior than would be possible without AI. The accounts could be used not only to propagate outright misinformation, but also to amplify negative but accurate information about a political candidate, thereby giving it more visibility among the electorate than it would have received absent the foreign influence.

A foreign government seeking to tip the scales in an election would have a long list of options for specific ways of undermining information integrity. A 2019 RAND Corporation report on "Hostile Social Manipulation" identifies over a dozen methods of social manipulation, including "content creation", "disinformation", "social media commenting", "direct advertising", "trolling", "behavioral redirection" and "microtargeting"[15]. With AI, all of these methods could be used at scale and in ways that might be difficult to mitigate, particularly given the importance of minimizing false positives, which could lead to suppression of legitimate social media content posted by real voters.

While election interference is an extremely important way in which nation-state might seek to use AI-generated misinformation to further geopolitical goals, it is not the only one. Nation-states might also use AI to disseminate information aimed at influencing a foreign government's geopolitically-relevant legislation; regulations; trade, economic, and defense policies; and decisions regarding major mergers and acquisitions. A nation state might also manipulate information to boost positive consumer perceptions of companies headquartered within the nation-state, thereby boosting the global competitiveness of those companies, and by extension, the nation-state. And, AI-enabled information manipulation will be a central feature of any future large-scale military conflict. This would include not only attempts to shape public opinion, but also efforts to undermine the availability and accuracy of information relied upon by military decisionmakers and political leaders.

Conclusion

So how can societies – and in particular democracies built on the free flow of information and ideas – address AI-enabled misinformation created and/or propagated by a foreign government?

[15] M.J. Mazarr et al., Hostile Social Manipulation, RAND, 2019.

Technology, policies, and awareness can all contribute to a solution. With respect to technology, as noted above, the same advances in AI that are making it easier to generate misinformation can also be used to detect it. Many of the tradeoffs involved parallel those found in cybersecurity, where there are also complex decisions to be made regarding how to allocate resources in relation to prevention, detection, and mitigation. The experience from that sector can help inform both public and private sector approaches to ensuring information integrity.

Governments should be both investing directly in research on improved detection as well serving as a resource for the private sector through information-sharing arrangements that can help companies better understand potential foreign manipulation of social media and other online information. The information flow can work in the other direction as well: Companies, and in particular social media companies, will be at the front lines of foreign-directed misinformation campaigns, and thus are well positioned to understand their dynamics and convey the lessons learned on to other companies and to the government.

Policy solutions can include the use of existing legal frameworks as well as new legislation. In considering the legal landscape, it is important to keep in mind that not all approaches that undermine information integrity will involve false statements. A foreign government might simply seek to amplify or suppress accurate information in ways aimed at swaying public opinion. When this occurs in the context of an election, it can be addressed through statutes aimed at combating election meddling. As important as such statutes are, their effectiveness will be limited due to the time scales involved (in many cases, the election will be long over by the time the legal system swings into action) and due to the fact that elections represent only one of the many potential targets of a misinformation campaign.

That highlights the importance of a final tool: increased awareness. In an era where deepfakes and other forms of manufactured or manipulated content will become more common,

broader awareness can help slow (though certainly not stop) their spread. In promoting this greater understanding, it will also be important not to undermine the trust in legitimate information which is at the foundation of all democratic societies.