AI and the 2024 election cycle

Digital Intelligence

BAE SYSTEMS

Since the widespread rollout of Artificial Intelligence (AI) powered chatbots and other tooling in 2022, the scope of AI/ML capabilities and applications has increased. This applies as much to malicious as legitimate use cases, with some citing concerns that AI tooling has reduced the 'barrier to entry' for threat actors, and could invite new AI-enabled attack techniques.

In the context of elections, the acceleration in quality, accessibility and application of generative AI (GenAI) makes it important to understand and prepare for its misuse. However, this is the first major election cycle since GenAI began to proliferate in this way, meaning the ways in which threat actors may use AI tools for interference purposes is unclear.

In this report, we take a closer look at AI risks to the 2024 election cycle, outlining current and possible applications of AI technologies and assessing their possible impact. The ultimate question is: do GenAI capabilities introduce new risks for elections in 2024, or do they just amplify existing risks to election infrastructure and the electorate?

Assessing the threat

From a malicious activity perspective, AI tooling offers three core advantages to attackers:

- **Speed:** AI tools increase the speed at which operational activities can be carried out
- **Scalability:** AI-driven malicious activities can be more easily amplified and operate at scale, possibly increasing their impact
- **Accessibility:** AI tools are becoming more sophisticated and available, lowering the barrier to entry for attackers

Out of all AI capabilities, GenAI tools are particularly relevant with regards to election interference. GenAI describes AI that can generate synthetic content, such as text, images or video. They are the most widely used form of AI, as they are scalable, accessible and support various business and individual use cases.

Public assessments argue that GenAI capabilities will likely not introduce new risks for the 2024 election cycle. AI has been used in some malicious operations, but these activities are unlikely to drastically impact voter preferences or alter vote counts. However, AI may amplify existing risks to election infrastructure and voters, meaning it is important to understand how these capabilities could impact the security and integrity of an election cycle.

Our assessment aligns with this view. AI technologies are unlikely to present a standalone threat to elections in 2024 and mainly confer tactical benefits to threat actors, which in an election context are most likely to impact the following areas:

- **Improved social engineering** through the creation of synthetic/deepfake content for lures and information operations
- **Enhanced cyber operations,** most notably reconnaissance phase techniques. This includes identifying misconfigurations, mapping out target networks or acquiring and processing personal information to use for social engineering. It may additionally enable more effective exploitation of vulnerabilities
- **Scaling and increased speed** of adversary campaigns, particularly those using social media
- **Somewhat reduced barrier to entry** for some actors, although the applications by low-skill actors are low impact
- **Attacks against AI systems** themselves

Let's dive into some of these areas and their potential impact.

Improved social engineering

The greatest concerns surrounding GenAI is around synthetic content creation and its impact on voters. As AI-generated content has proliferated, distinguishing between factual and false information has become increasingly difficult. This is especially true when combined with disinformation campaigns that mislead voters in order to influence their decision making.

Synthetic content creation appears to be the predominant use case for GenAI in election influence. Numerous countries with elections this year have reported at least one political deepfake in 2024, and several have warned that "various threat actors will likely attempt to use AI-produced content to influence and sow discord". So far, it has been observed in the following contexts:

- **Used by political candidates** to support their campaigning, including messaging and to promote/discourage voter turnout
- **Used by various actors** to aid attempts to discredit political candidates, or spread conspiracies around administrations
- **Used to promote and/or endorse** scams and fraudulent services
- **Used to undermine** public trust in democratic processes themselves
- **Used to tactically aggravate** and politicise emergent events or issues

In the context of elections, the proliferation of inauthentic, misleading material risks degrading the information environment — causing a significant erosion in public trust and making governance more challenging.

However, assessing the impact of disinformation and deepfakes isn't straightforward: it depends on measuring changes in voter behaviour. With the quality of 'deepfakes' constantly improving, the 'believability' of synthetic content may increase throughout 2024 and beyond. Believability also depends on the context in which the deepfake is used; for example, voters may be more likely to believe content if it is deployed at speed or is combined with some factual information. Recent examples include:

In January 2024, voters in New Hampshire received voice calls allegedly from President Biden encouraging them not to cast their vote in the New Hampshire Presidential Primary Election. This was one of the first instances of voice-based deepfakes being used at scale in order to deter voters from participation. The case was linked to a US political campaign group that has received a $6 million fine and numerous criminal charges.

In South Korea, a deepfake video of President Yoon admitting to corruption went 'viral' and was amplified in mainstream media ahead of the 2024 elections. South Korea's National Election Commission later reported that the video was to be taken offline. Within one month, the agency reported 129 election-related deepfakes.

In Ireland, deepfake pornographic content of politicians and public figures has been circulated in order to discredit and intimidate them. Women in public-facing roles have been particularly impacted by this trend.

To complicate this, 2024 has seen an uptick in political campaigns using AI-generated content to bolster their cause. From a voter perspective, this adds to the blurring of boundaries between legitimate campaign content and disinformation, making it harder to identify disinformation. This may also incentivise the growth of deepfake and AI-as-a-service offerings in legitimate and criminal marketplaces. As AI-generated content becomes normalised, voters may become more suspicious of authentic information (some of which may also be AI-generated), particularly during crises or political conflicts.

Enhanced cyber operations

Several threat actors have been observed using AI in their general cyber operations. However, at the time of writing there have been no reports of actors using AI in cyber operations targeting election infrastructure specifically.

Most obviously, AI technologies can enhance reconnaissance-stage techniques. This includes improved information gathering to developing social engineering materials, as well as vulnerability scanning. In 2024, the FBI warned that: "Cybercriminals are leveraging publicly available and custom-made AI tools to orchestrate highly targeted phishing campaigns, exploiting the trust of individuals and organizations alike."

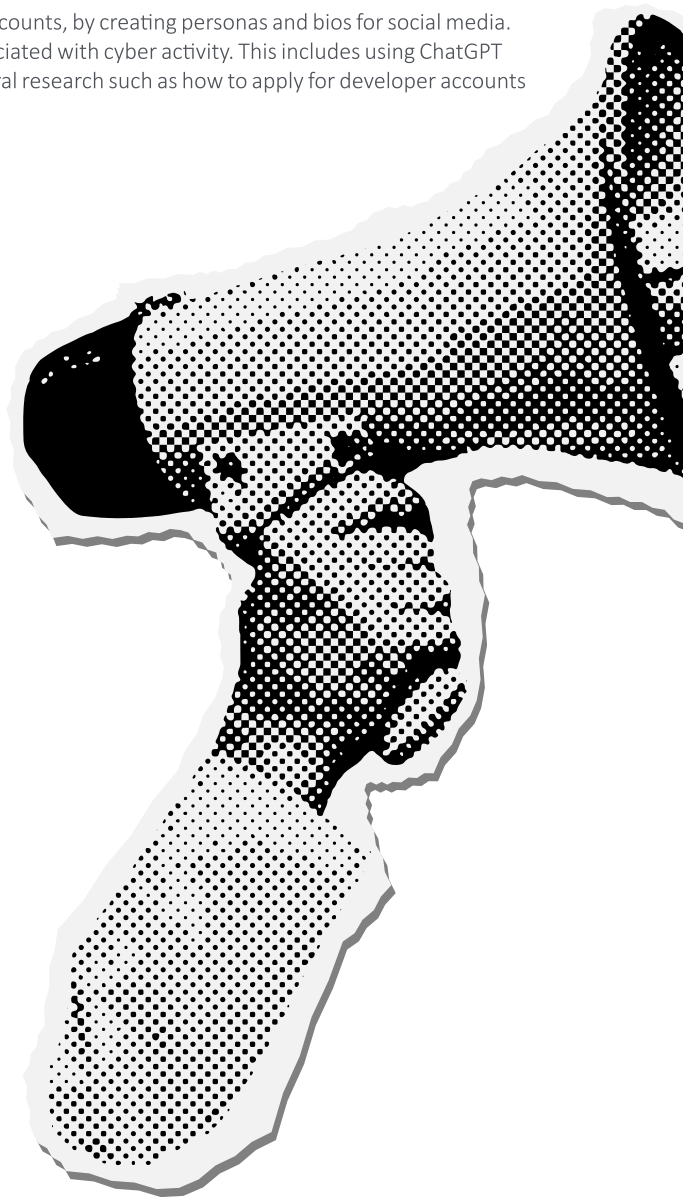
Chatbots such as ChatGPT have been used to develop more authentic sockpuppet accounts, by creating personas and bios for social media. AI has also been described as removing the operational burden, or 'grunt work', associated with cyber activity. This includes using ChatGPT to debug code for malicious websites and using the same tool to conduct more general research such as how to apply for developer accounts on social media or asking for summaries of social media posts.

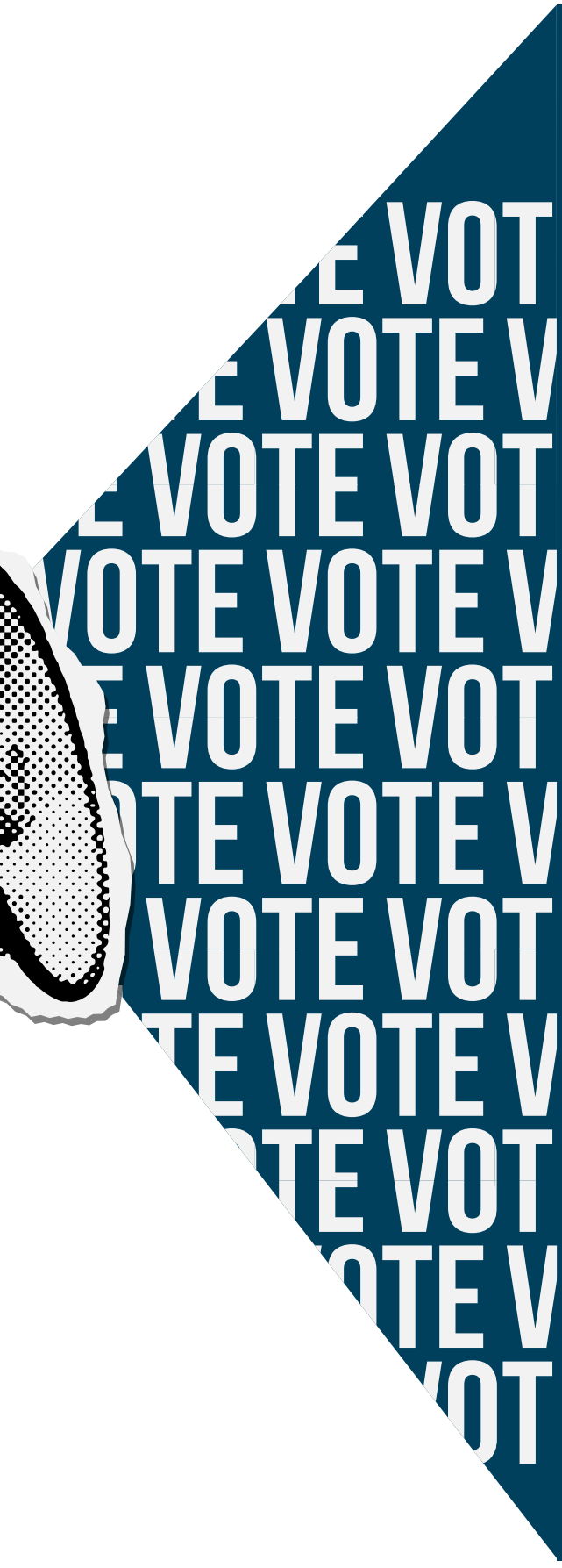
Attacks on AI systems

GenAI ecosystems themselves may offer an attack vector for threat actors. In 2024, two techniques have been particularly cited with reference to election security.

The first of these is Prompt Injection. This involves threat actors crafting malicious prompts that make Large Language Models (LLMs) act in unintended ways – such as leaking sensitive data or spreading disinformation – and includes vulnerabilities in widely used chatbots. For example, in 2023 a prompt injection vulnerability was identified in ChatGPT, which allowed attackers to feed malicious websites through ChatGPT plugins and exfiltrate conversation history.

The second technique is Data Poisoning, which involves attackers compromising an LLM's training dataset in order to influence or manipulate its operation. These activities can introduce bias, create erroneous outputs, introduce vulnerabilities, and influence the decision-making or predictive capabilities of a model in other ways. No data poisoning attempts have been observed in relation to an election, however there are indications that manipulation of datasets and prompts are being used to implement censorship.





What's the verdict?

When it comes to elections, AI is most likely to present a threat in three areas – **disinformation campaigns, some components of cyber operations, and the security of AI itself.** However, in all of these areas the gains are mainly tactical, meaning the speed and scalability of existing techniques may be 'AI-enhanced'.

Standard cyber security advice applies to these techniques, and defenders will be able to mitigate many of the risks AI poses through good cyber hygiene practises, awareness training and technical mitigations. Much of this guidance is publicly available – such as here and here.

It should be cautioned that AI is a tool, and the threat it poses to elections depends on which threat actors are using it and for what purpose. Evidence of AI use in connection with election interference from 'high-end' actors remains sparse, with current evidence pointing to rudimentary applications by lower-capability threat actors.

Nonetheless, it is unclear how the threat AI poses to elections and in the coming years democracies more generally will evolve in coming years. In this context, it is important to highlight that the problem is more complex than any single organisation is capable of solving in isolation, and effective collaboration between private industry, government, media and the wider public will be required to counter it.

We are Digital Intelligence

BAE Systems Digital Intelligence is home to 4,700 digital, cyber and intelligence experts. We work collaboratively across 16 countries to collect, connect and understand complex data, so that governments, nation states, armed forces and commercial businesses can unlock digital advantage in the most demanding environments. Launched in 2022, Digital Intelligence is part of BAE Systems, and has a rich heritage in helping to defend nations and businesses around the world from advanced threats.

BAE Systems Digital Intelligence
Surrey Research Park
Guildford
Surrey GU2 7RQ
United Kingdom
T: +44 (0) 330 158 3627

BAE Systems Digital Intelligence
Malta Office Park
ul. Abpa A. Baraniaka 88
Poznan
61-131
Poland
T: +44 (0) 330 158 3627

BAE Systems Digital Intelligence
Level 2
14 Childers St
Canberra
ACT 2601
Australia
T: +61 (0) 2 9053 9330

BAE Systems Digital Intelligence
Level 15-19
550 Bourke Street
Melbourne Vic 3000
T: +61 884 80 8888

BAE Systems Digital Intelligence
Level 28, Menara Binjai
2 Jalan Binjai
Kuala Lumpur
50450
Malaysia
T: +60 327 309 390

BAE Systems, Surrey Research Park, Guildford, Surrey, GU2 7RQ, UK

E: learn@baesystems.com

W: baesystems.com/digital

linkedin.com/company/baesystemsdigital

@BAESystemsDigi

BAE SYSTEMS