Cyber security guidance for democratic institutions: artificial intelligence

CANADIAN CENTRE FOR CYBER SECURITY

June 2024 | ITSAP.00.135

Artificial intelligence (AI) is a continuously developing technology that uses computer algorithms to performs tasks, predict results, and create content. Generative AI is a specific subset of AI that uses large data sets to create new content. It is without question that AI has complicated the democratic landscape. The increased availability of large language models (LLMs), chatbots, and synthetic media-creation technology has made it easier for threat actors to disrupt democratic campaigns and elections. Cyber threat actors can create and disseminate manufactured content to cause confusion, discredit the electoral system and spread a false narrative to sway voters. They can use AI to increase the output of distributed-denial-of-service (DDoS) attacks or create malware that is undetectable by regular anti-virus and malware protections. They can also use LLMs to produce and widely disseminate highly personalized and realistic communication. LLMs are inexpensive and easily accessible to anyone wishing to disrupt an election campaign. Because of these capacities, people working for democratic institutions must be hypervigilant when using email, social media and other communication channels. AI also enhances data collection. In the hands of threat actors this can be used to quickly collect and distribute public information. This can then be used against candidates and staff to carry out activities like doxing attacks. Doxing is the act of providing personal information online, such as addresses, typically with malicious intent. AI technologies are rapidly evolving and the mitigation strategies in this publication may not be enough protection in the future. Additional mitigations and updates to guidance will continue to be provided as AI evolves.

Types of fake or synthetic content

- **Deepfakes**: Images or recordings that have been altered to show someone saying or doing something controversial or sensational. They are generally convincing in their appearance and made with the intent to deceive.
- **Fake social media accounts**: Used to impersonate politicians or those in the community who have influence, such as celebrities.
- **Voice cloning**: Use of samples of speeches, interviews, and various other audio clips to impersonate someone. Voice cloning can be used to impersonate election authorities and staff to gain access to sensitive information, to spread disinformation or to discredit a candidate.
- **Disinformation**: Used to provide false information on topics related to elections, such as where and when to vote, with the aim of disrupting the voting process. The intention is to cause confusion and lower voter turn out in a targeted population.

Artificial intelligence bias and training

LLMs are algorithms used by programs that can understand prompts and generate human-like responses. They are trained on large datasets to analyze, summarize, translate and generate content. However, their output is only as great and as accurate as their input.

These programs rely on available online content. Any information that is used as inputs or prompts may also be used for training the models. You must be aware of what types of content and guardrails LLMs are receiving, as well as what kinds of content are not included or available. There may be an overwhelming influx of data from a certain part of the world or a specific demographic group, while other areas and groups may only be represented by a small sample of data or none at all. This discrepancy is called bias. It's impossible to completely remove it from LLMs, but we must be aware of it.

Underrepresented groups, such as ethnic or linguistic minorities, may be unfairly represented or not represented at all in LLMs' output. Excluding certain demographic groups prevents LLMs from engaging in accurate political discussions.

How to mitigate the threat of artificial intelligence

- Protect sensitive information by implementing access controls, strong passwords and multi-factor authentication (MFA)
- Implement security controls on staff email, media platforms and other work devices
- Harden your organization's social media accounts by:
  - deactivating or deleting profiles that are no longer in use
  - removing any personal identifiable information
  - making personal profiles private
- Develop zero trust policies to limit access to certain information and applications
- Take extra time and care to verify information before responding to it
  - Be selective when engaging in online commentary
- Adopt a policy of a rolling passphrase, that only authorized personnel know, to prevent adversaries from gaining access to organizational systems via real-time voice cloning
  - A rolling passphrase or password is often automatically generated and changed at regular intervals to deter unauthorized access
- Educate staff members, your followers and the public about:
  - the potential risks and vulnerabilities of using and consuming media
  - media literacy
  - what to watch for online
  - how to identify reliable sources and verify information
- Report incidents of deepfakes to the platform on which they are hosted and the Canadian Anti-Fraud Centre (CAFC) using their Online Reporting system or by phone at 1-888-495-8501
- Proactively draft a Frequently Asked Questions page or a set response for inquiries about synthetic content
- Ensure networks that host electoral processes are designed and deployed with strong authentication mechanisms, such as MFA
- Develop an incident management plan

Related resources

- Cyber Threats to Canada's Democratic Process: 2023 Update
- Engaging with artificial intelligence
- Guidelines for secure AI system development
- Steps for effectively deploying multi-factor authentication (MFA) (ITSAP.00.105)
- An Artificial Intelligence strategy for NATO
- The threat from large language model text generators
- Risk in focus: Generative A.I. and the 2023 election cycle

Communications Security Establishment Canada | Centre de la sécurité des télécommunications Canada

© His Majesty the King in Right of Canada, as represented by the Minister of National Defence, [2024]

Canada