ARTIFICIAL INTELLIGENCE

AUGUST 2020

ITSAP.00.040

The world we live in is being transformed by artificial intelligence (AI). For example, when you use facial recognition to access your mobile device or ask your smart speaker for the weather forecast, it is AI that makes it possible. Simply put, this developing technology uses intelligent computer programs (i.e. learning algorithms) that find complex patterns in data to solve a problem. The data allows an AI machine learning tool to behave in a way that would be considered intelligent if done by a human (e.g. understanding languages or learning from experience).

WHAT IS AI GOOD AT TODAY?

Machine learning tools are good at solving problems where the solution is found in the data provided. It is important for the tool to be trained using proper and accurate data to get the best solutions. Machine learning tools continue to learn as you provide them with more data and feedback, eliminating the need to replace the tool.

WHAT IS AI NOT SO GOOD AT TODAY?

Machine learning tools are not so good at solving problems where reasoning or common sense (i.e. additional context) is required. Often, it is too expensive and challenging to provide a machine learning tool with all the data necessary to allow it to operate with 'common sense'. Humans, with their judgement and insight, are still better able to handle these types of decision-making situations.

WHAT ARE SOME OF THE WAYS IN WHICH ORGANIZATIONS ARE USING AI?

Facial recognition: AI (specifically an area called neural networks) has improved facial recognition solutions by making them faster and more accurate.

Process optimization: A properly trained machine learning tool (one learning from accurate data) can use the data to give more accurate solutions and perform mundane tasks faster than a human can.

Chatbot: These common machine learning tools use natural language processing to automate customer service chats. A customer can ask a chatbot a question and get a response within seconds—24 hours a day, seven days a week.

Computer-assisted diagnostics: Machine learning tools are used in the medical industry to aid in patient diagnosis. The tools learn from previous cases and provide a second opinion based on what it has learned.

Fraud detection: Sophisticated machine learning tools can detect fraudulent emails faster than a human can. These tools sort through your inbox and move spam and phishing emails to your junk folder.

Call centre assistance: Machine learning tools may change the way call centres function. With AI, wait times might be reduced and call centre employees could be reallocated to other, less stressful positions.

AWARENESS SERIES

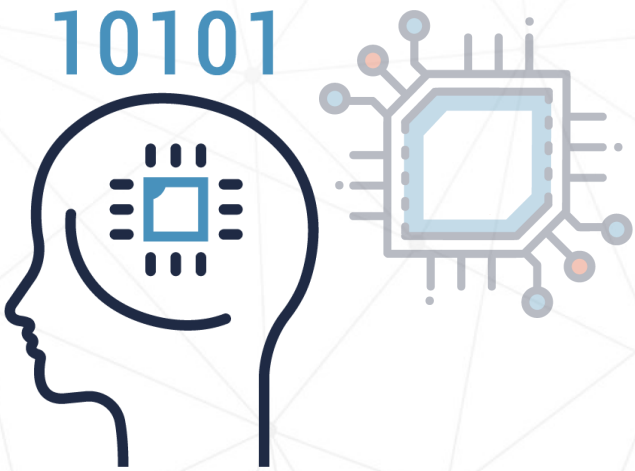
WHAT ARE THE THREATS TO AI TOOLS?

AI tools are often only as good as the data model they rely upon. The main threats to AI come from compromises to its data. Common methods of compromise include:

Data Poisoning Attack: This type of attack occurs at a machine learning tool's training phase. AI tools rely heavily on accurate data for training. When poisoned (inaccurate) data is injected into the training data set, the poisoned data can lead the learning system to make mistakes.

Adversarial Example: This type of attack occurs after the machine learning tool is trained. The tool is fooled into classifying inputs incorrectly. For example, in the case of autonomous vehicles, an adversarial example could be a slight modification of traffic signs in the physical world (subtle fading or stickers applied to a stop sign), causing the vehicle's AI system to misclassify a stop sign as a speed-limit sign. This could seriously impact the safe operation of self-driving vehicles.

Model Inversion and Membership Inference Attacks: both of these scenarios occur when a threat actor queries your organization's data model. A model inversion attack will reveal the underlying data set, allowing the threat actor to reproduce the training data. A membership inference attack confirms if a specific data file is part of the training data. Both model inversion and membership inference attacks could compromise the confidentiality and privacy of your training data and expose sensitive information.



WHAT ELSE SHOULD YOU KNOW ABOUT AI?

- Machine learning tools can detect patterns in data.
- Machine learning tools need enough data to see the patterns at a high enough frequency.
- Data used for training should be complete, diverse, and accurate.
    - If there are blanks in the data, some patterns might not be discovered, and the patterns that are found might not be accurate.
    - If the data used is not diverse, the tool will have a narrow scope.
    - If the training data used is not accurate, the tool will provide unreliable results.
- Data that is recorded and collected for "quality control" purposes can contain both sensitive and personal information.
- Many organizations are now using trustworthy AI policies to ensure that their use of AI tools minimize potential biases and unintended consequences, especially regarding the treatment of individuals. Policies may also assist in the development of appropriate protocols for the handling of sensitive and personal information An example of an AI policy is the Government of Canada's recently adopted Directive on Automated Decision-Making.
- If your organization intends to deploy AI, it should consider seeking legal advice to manage the many ethical, privacy, policy, and legal considerations that come from using AI.

Need help or have questions? Want to stay up to date and find out more on all things cyber security?
Visit the Cyber Centre website at **cyber.gc.ca**