Communications Security Establishment
Centre de la sécurité des télécommunications

CANADIAN CENTRE FOR CYBER SECURITY

Artificial Intelligence

July 2023

ITSAP.00.040

The world we live in is being transformed by artificial intelligence (AI). This developing technology uses intelligent computer programs (i.e. learning algorithms) to find complex patterns in data to make predictions or classifications. AI is used today to perform specific tasks, such as to use facial recognition to access your mobile device or ask your smart speaker for the weather forecast. Machine learning, a subset of artificial intelligence, uses instructions, known as algorithms, and data to understand languages to help the computer system learn and improve based on its own experience. Deep learning, a subset of machine learning, uses vast volumes of data and artificial neural network algorithms to train a model to make intelligent decisions on its own.

What can AI do?

AI already plays a big role in our everyday lives. From search engines to online shopping to voice assistants on our mobile device or smart speaker, AI provides recommendations, information, answers to questions, and helps to organize our schedules. These daily applications create data and feedback for machine learning tools to learn and improve from.

What can't AI do today?

A few fundamental limitations still exists for AI today. Using reasoning or common sense, and adapting to different situations, and understanding cause and effect are all quite difficult for AI. Humans, with their judgement and insight, are still better able to handle situations that require these types of problem solving and decision making skills.

What are some of the ways in which organizations are using AI?

Facial recognition: A leading application of AI that looks at facial features in an image or video to identify or verify the individual.

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.

Digital assistants: Chat or voice bots can improve customer service and reduce support costs. Customers can receive help within seconds—24 hours a day, seven days a week. These services are often highly personalized and can be based on a user's preferences and history with the organization.

Healthcare: In the medical industry, AI aids in patient diagnosis and treatment in a variety of ways, such as computer-aided diagnostic systems that assist in making a diagnosis. Machine learning in precision medicine is another highly useful tool and is used to help predict which treatments are most likely to succeed on a patient.

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.

Data analysis: Using machine learning algorithms, AI is capable of analyzing large amounts of data and discover new patterns. This greatly reduces the processing time spent by a data analyst, known as automation, and improves business performance.

Cyber security: AI is useful in detecting new threats to organizations through automation. By using sophisticated algorithms, AI are able to automate threat detection such as malware, run pattern recognition to find relationships between different attack vectors and provides superior predictive intelligence.

AWARENESS SERIES

Cat. No. D97-1/00-040-2022E-PDF
ISBN 978-0-660-44248-8

Canada

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.

Generative artificial intelligence

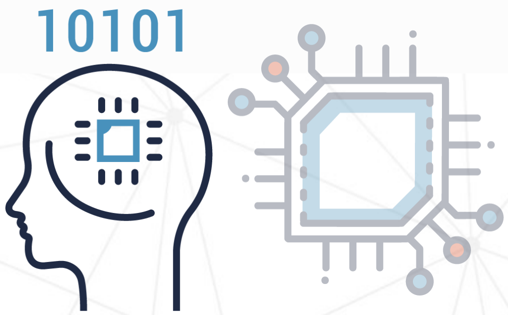
This is a type of artificial intelligence that generates new content by modelling features of data from large datasets that were fed into the model. While traditional AI systems can recognize patterns or classify existing content, generative AI can create new content in many forms, including text, image, audio, or software code.

One class of generative AIs that have seen significant improvement in recent years are large language models (LLMs). Since late 2022, several LLMs (OpenAI's ChatGPT and Google's LaMDA) and services using LLMs (Google's Bard and Microsoft's Bing) have gained the world's attention.

For more information, see Generative artificial intelligence (ITSAP.00.041)

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**