Engineering, Operations & Technology Committee

Introduction to AI

Item 6c
July 8, 2024

Item 6c
Introduction to AI

Subject
Introduction to AI

Purpose
This item provides an overview of the different types of Artificial Intelligence (AI) and the steps IT is taking to prepare for the adoption of AI tools.

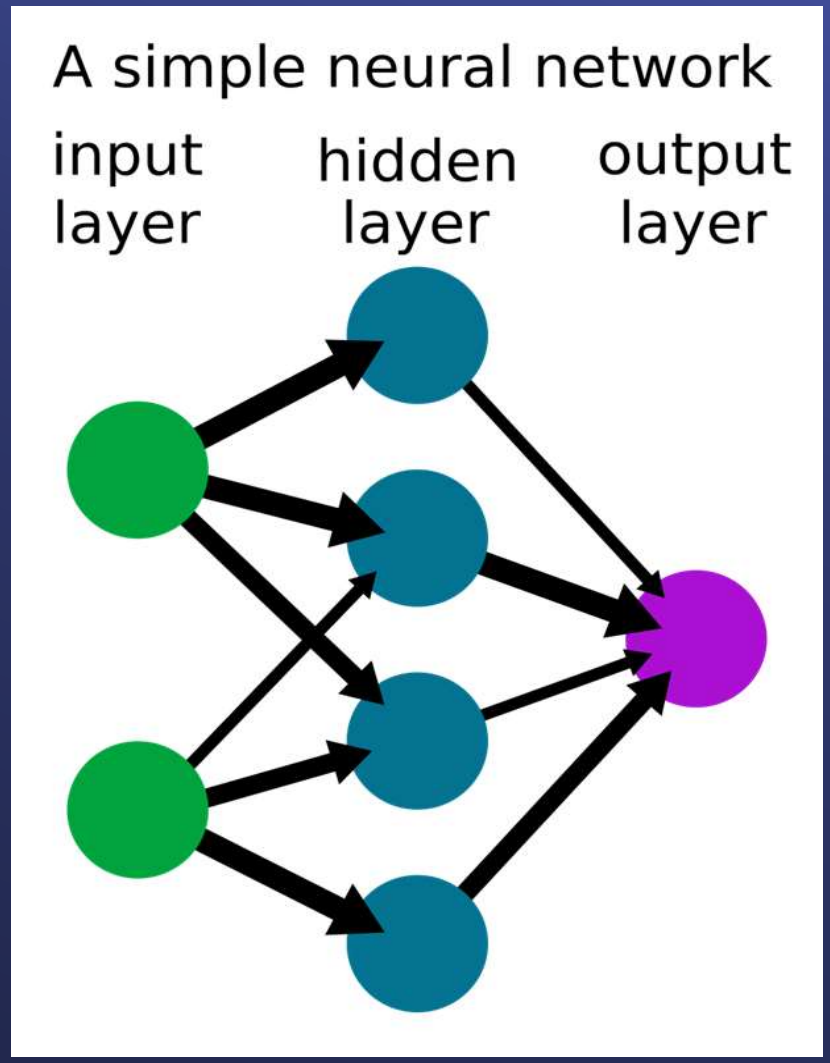
Agenda

In this presentation

- How AI Works
- Recent history of AI
- Governance and Risk
- Capabilities, and Opportunities at Metropolitan

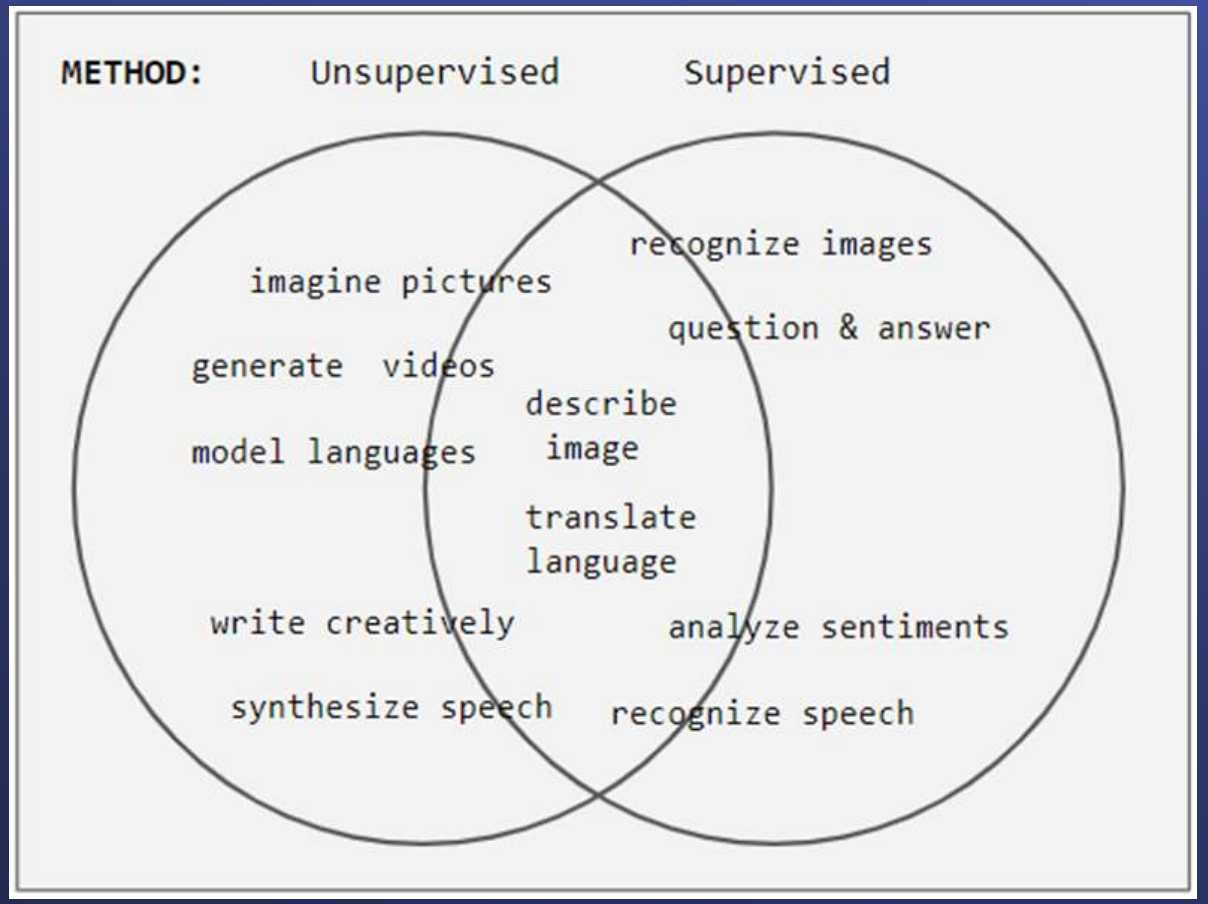
How AI Works

- Neural Network model inspired by the structure of animal brains
  - Artificial neurons are connected by “edges” which model a brain’s synapses
- Neuron receives signals from the connected neurons, and produces an output based on its activation function
- Neurons are aggregated into layers which each perform different transformations



How AI Works

- There are many methods for training artificial neural networks:
  - Supervised learning (spelling bee)
  - Reinforcement learning (rats with a cheese button)
  - Semi-supervised learning (example problems at beginning of an exam)
  - Unsupervised learning (finding patterns)



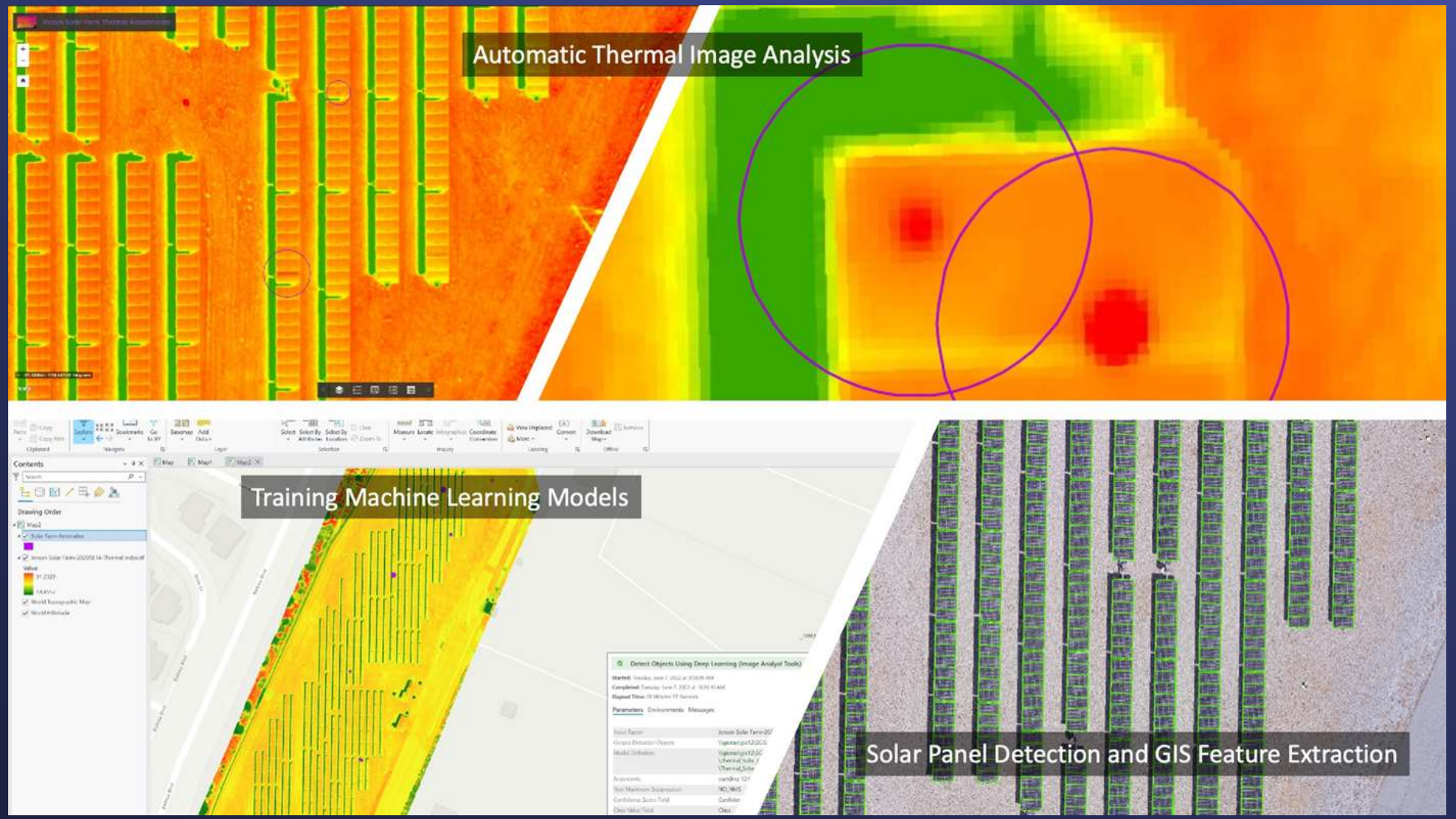
Recent History

- The term “Machine Learning” was coined in 1959 by Arthur Samuel of IBM
- Advances in the 90’s brought us data-driven Machine Learning
- Advances in the 2000’s and 2010’s brought us Deep Learning, which is capable of unsupervised learning
  - The term “deep” refers to the number of layers of nodes
  - “Unsupervised learning” means the AI can determine, without being told by a human, which attributes are most important to consider for an accurate outcome

Recent History

- Popular commercial applications of Machine Learning or Predictive AI in the past few decades include…
  - Siri, Roombas, Ring, and Fraud Detection

2022 GovX Project Experience Winner: MWD!



Recent History

In 2014, the development of the generative adversarial network at the University of Montreal enabled deep neural networks to capture an aesthetic by using an opposing “generator” and “discriminator” to determine its own successes. This led to modern AI art generators.

Source: Bing image creator
Prompt: “an oil painting of the Hoover dam”

Recent History

- In November of 2022, OpenAI's ChatGPT was released for public testing, which was the first broad exposure to a Large Language Model (LLM).

A Word on ChatGPT and LLM Limitations

- It's designed to interface with humans.
  - Limits its usefulness in things like generating database queries without fine-tuning.
- Non-deterministic nature makes it feel more human but means results will not be consistent.
  - Consistency is crucial for integration with applications – computers can't understand English.

A Word on ChatGPT and LLM Limitations

- ChatGPT failed a 6th grade Singapore math test in February 2023.
- Trained explicitly on language.
  - Not ideal for something like math, with infinite possible permutations.
  - May be accidentally correct by having the problem in its training data but is not capable of multistep logical inference.
- Can 'hallucinate', or return incorrect results with a high degree of confidence.

Source: Bing image creator
Prompt: "a series of road signs"

Governance and Risk

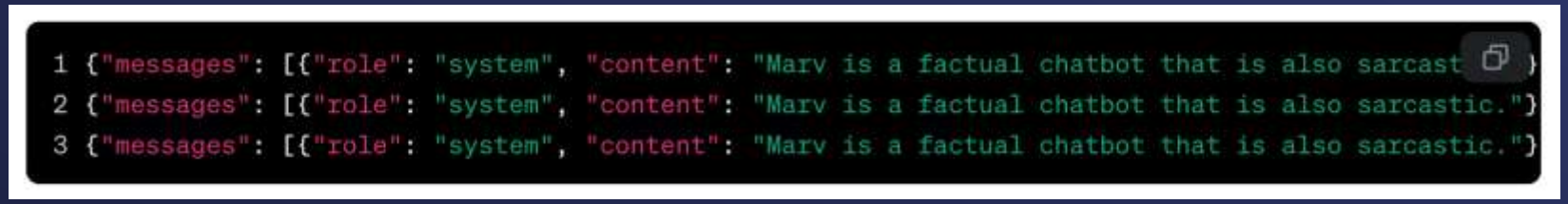
- Metropolitan published our GenAI Guidelines in March 2024.
  - Include a disclaimer when sharing AI-generated content.
  - Do not use sensitive data as an input to any AI tool.
  - Fact check all outputs.
  - Enterprise AI tools must be vetted by IT.
  - We are accountable for what AI creates at our direction.
  - Do not appropriate content for use in AI tools without permission, such as artwork.
  - Managers should understand the guidelines and assist staff accordingly.

Governance and Risk

- Data classification is required to safely deploy AI.
- The number one concern identified by a Gartner survey of over 600 technology leaders when asked about AI was about privacy of data.
  - Many AI tools “scrape” your enterprise data.
  - Improperly classified data can be unintentionally exposed by AI.
  - With the advent of ”Action Models”, improper permissions and classifications may lead to AI taking actions that harm the enterprise.

Governance and Risk

- Fine tuned models.
  - Enterprise LLM solutions are typically priced by input/output volume.
  - Fine tuning provides control over the nature of responses the LLM provides.
  - Allows you to add unseen parameters to a user's input to steer the LLM generation toward a specific purpose.



What's Been Done to Date

- Many users throughout Metropolitan have evaluated ChatGPT and other GenAI models such as DALL-E for image creation.
- Metropolitan IT is working with Microsoft to perform an "AI Readiness" check based on our current infrastructure and data security posture.
- Machine learning has already been used successfully at Metropolitan for analytical work, cybersecurity, and performance monitoring tools.

Next Steps

- Implement new data classification tools through the Enterprise Content Management project.
- Continue testing of GenAI tools for code generation, database querying, and automating repeatable processes.
- Refining our standards around AI to ensure responsible and ethical use of the technology.
- Microsoft Office Copilot testing and deployment (draft documents, perform analysis, create presentations).
- Model transition from prompt engineering to leading questions.
- GenAI integration with other emerging technology domains, such as augmented reality and 3D printing.


THE METROPOLITAN WATER DISTRICT OF
SOUTHERN CALIFORNIA
1928