Canadian Food Inspection Agency | Agence canadienne d'inspection des aliments

Foundational AI concepts

Presented by: **Noureddine Meraihi**
Manager, AI-Lab, Digital Services Branch, CFIA

For: 48 th North American Plant Protection Organization (NAPPO) Annual Meeting

Canada

TABLE OF CONTENTS

What Is Artificial Intelligence?

Despite over 70 years of research, there's still **no single universally accepted definition** of AI. Different organizations emphasize different aspects:

Canadian Government Definition

"A computer system with the ability to complete complex tasks on its own by recognizing and replicating patterns identified in data."

—Innovation, Science and Economic Development Canada

OECD Definition

"A machine-based system that, for explicit or implicit objectives, infers from the input it receives how to generate outputs—such as predictions, content, recommendations, or decisions—that can influence physical or virtual environments."

—G7 Toolkit for Artificial Intelligence in the Public Sector, 2025

Both definitions highlight AI's ability to **learn from data**, **recognize patterns**, and **make autonomous decisions**—capabilities that fundamentally distinguish it from traditional software.

The AI Lifecycle

Data collection

Gathering, cleaning, labeling, and transforming raw information into usable formats

Model design & training

Selecting architectures and teaching the system to recognize patterns.

Evaluation & Validation

Testing performance against benchmarks and real-world scenarios

Deployment

Integrating into operations and continuously tracking behavior

Responsible AI Practices (FASTER)

Ensuring the AI is **F**air, **A**ccountable, **S**ecure, **T**ransparent, **E**ducated and **R**elevant

Essential Terminology

Data

Information used to train AI models and generate predictions—the foundation of all AI systems

Model

The mathematical engine that maps inputs to outputs through learned patterns

Training

Adjusting a model's internal parameters to minimize error between predictions and actual outcomes

Generative AI

Models that create new content (text, images, code) by learning from patterns in training data.

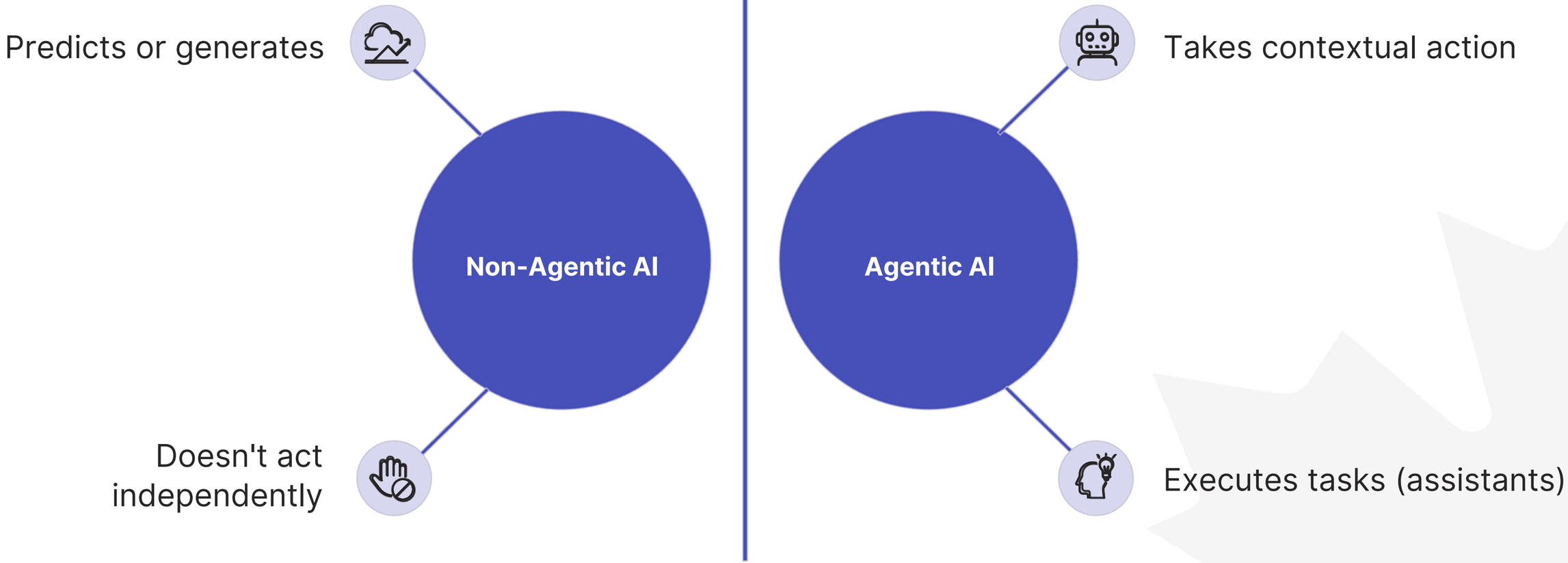
AI System

An integrated setup combining one or more models with interfaces, filters, and decision rules

AI Agent

General-purpose AI that can plan, adapt, and act toward goals while interacting with its environment

Agentic vs non-agentic



Data Collection & Preparation

Developers must collect, clean, label, and transform raw information into formats machines can learn from.

Data Types in Plant Protection

- Satellite and drone imagery
- Sensor readings (temperature, humidity)
- Genomic sequences
- Historical pest outbreak records
- Expert annotations and classifications
- Trade import data

Critical Quality Principles

- **Representativeness:** Coverage across regions, seasons, and conditions
- **Diversity:** Multiple crop varieties and pest species
- **Label accuracy:** Expert verification of classifications
- **Documentation:** Clear metadata and provenance tracking

Bias and fairness chanllenge

AI Bias -*Systematic errors or unfairness in the outcomes, predictions, or decisions made by artificial intelligence systems. These biases can arise from various stages of an AI system's lifecycle*

Data collection

Model design & training

Deployment

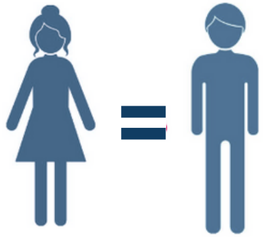
AI bias often reflects or amplifies biases present in the data, processes, or assumptions used to build the system. It can lead to unequal or discriminatory outcomes, affecting ***fairness****,* ***accuracy****, and trustworthiness of an AI system.*
(Accessibility Standards Canada, 2025)

Fairness Metrics

Individual fairness is based on the idea that similar subjects with respect to the predicted task should be treated similarly. It assumes that there exists a metric that allows defining similarity between individuals.

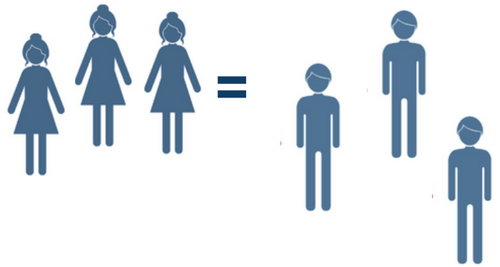
This is called the Lipschitz condition:

> Two individuals x and y, at a distance $d(x,y) \in [0,1]$, are respectively associated with distributions M(x) and M(y) such that the statistical distance between M(x) and M(y) is at most d(x,y).



Zhou et al., "Bias, Fairness, and Accountability with AI and ML Algorithms"

Group Fairness



Demographic parity

A system is said to satisfy demographic parity when all prediction P(Y = 1) is independent of the sensitive class or the rate at which all groups are classified as positive is the same for all groups. This is also called the "statistical parity" criterion.

$$P[\hat{Y} = j \mid A = 0] = P[\hat{Y} = j \mid A = 1], j \in \{0, 1\}$$

Example: Men and women have the same probability of being classified by the same category.

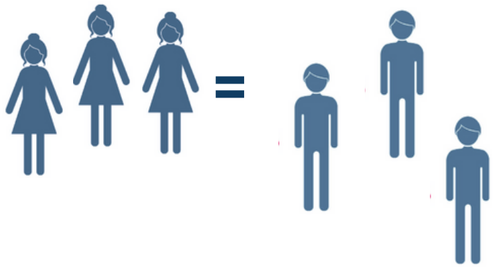
Conditional Demographic Parity

An AI system respects conditional demographic parity when for any prediction Y, the probability of belonging to the positive or negative class is the same for all groups A, while controlling for a set of legitimate attributes L.

$$P[\hat{Y} = j \mid L = l, A = 0] = P[\hat{Y} = j \mid L = l, A = 1], j \in \{0, 1\}$$

Example: Men and women with the same qualifications have the same probability of being hired by a company.

Group Fairness



Equalized Odds

An AI system respects equalized odds when all groups have the same **error rate.** This notion satisfies the separation: the probability of being correctly or incorrectly classified is the same.

$$P[\hat{Y} = j \mid A = 0, Y = k] = P[\hat{Y} = j \mid A = 1, Y = k], j, k \in \{0, 1\}$$

Example: Men and women have the same probability of being hired by a company, even if they are not qualified.

Predictive Parity

An AI system satisfies this definition if the error rate among people receiving the same decision is the same for all groups.

$$P[Y = k \mid \hat{Y} = j, A = 0] = P[Y = k \mid \hat{Y} = j, A = 1], k, j \in \{0, 1\}$$

Example: Men and women who were rejected by the system have the same qualifications.

Model Design & Training

During training, developers design sophisticated architectures and feed them diverse datasets to teach pattern recognition. The model learns by repeatedly adjusting billions of internal parameters.

Common Training Challenges

Overfitting

Memorizing training data instead of learning general patterns

Underfitting

Failing to capture important relationships in data

Bias propagation

Amplifying existing prejudices in training data

Explainability

The "black box" problem of understanding decisions

Critical Question: Do we truly understand how our model generalizes beyond training data? What hidden assumptions is it making?

Why Transparency Is Hard

Complexity

Deep models contain millions of parameters—hard to interpret.

Proprietary restrictions

Commercial models hide data and architecture details.

Non-determinism

Same input may produce different outputs due to randomness.

Opaque data origins

Limited visibility into training datasets.

Critical Question: Can you explain why ChatGPT generates one specific answer instead of another?

Deployment & Monitoring

Developers deploy the AI system for **real-world use**, collect feedback, monitor performance, and continuously improve it based on observed issues.

Deployment — From Model to System

Integrate trained models into real-world applications or services.

Continuous Safety Loop

Detect risks and performance issues once the system is live.

Responsible AI isn't a checkbox at the end—it must be **embedded throughout every stage** of development and deployment. Each phase introduces unique risks that require specific mitigation strategies.

Reflection questions and Q&A

Critical Question: Who is accountable when an AI system fails—the developer, the deployer, or the end user? How do we detect problems early enough to prevent harm?

Thank you