AI Cybersecurity

About Me

Shay Ya'ari

- Software Engineer at Botanica Consulting
- 15 years building software
- From Israel, based in Japan

Debug code for work, reality for fun.

Learning Objectives

- Describe the anatomy and functionality of AI systems and their relationship to attack surface and attack vectors
- Evaluate AI system proposals for security risks
- Develop control requirements, mitigations and security specifications for AI deployments
- Assess AI security incidents and determine appropriate responses
- Brief technical and non-technical stakeholders on AI security decisions

Schedule

Day	Time
Day 1 **The Anatomy of AI Systems**	**9:00 - 17:00**
Days 4 **Cybersecurity Attacks and Defenses of AI Systems**	Break **10:30 – 10:45**
Day 4 **AI Threat Modeling**	Lunch **12:30 - 13:30**
Day 5 **AI Incident Response & GRC**	Break **15:30 – 15:45**

Motivation

Nano Banana

Demo

Claude Artifact

Demo

NotebookLM

Demo

Veo 3.1

Demo

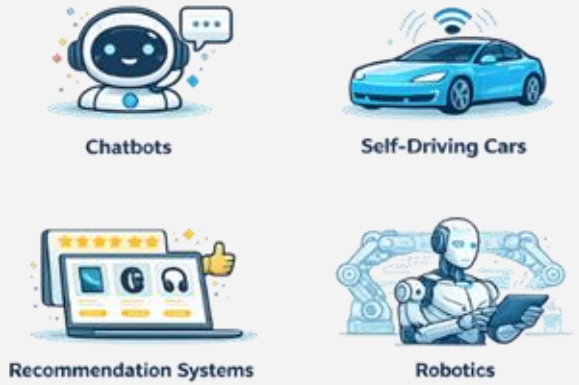
DART

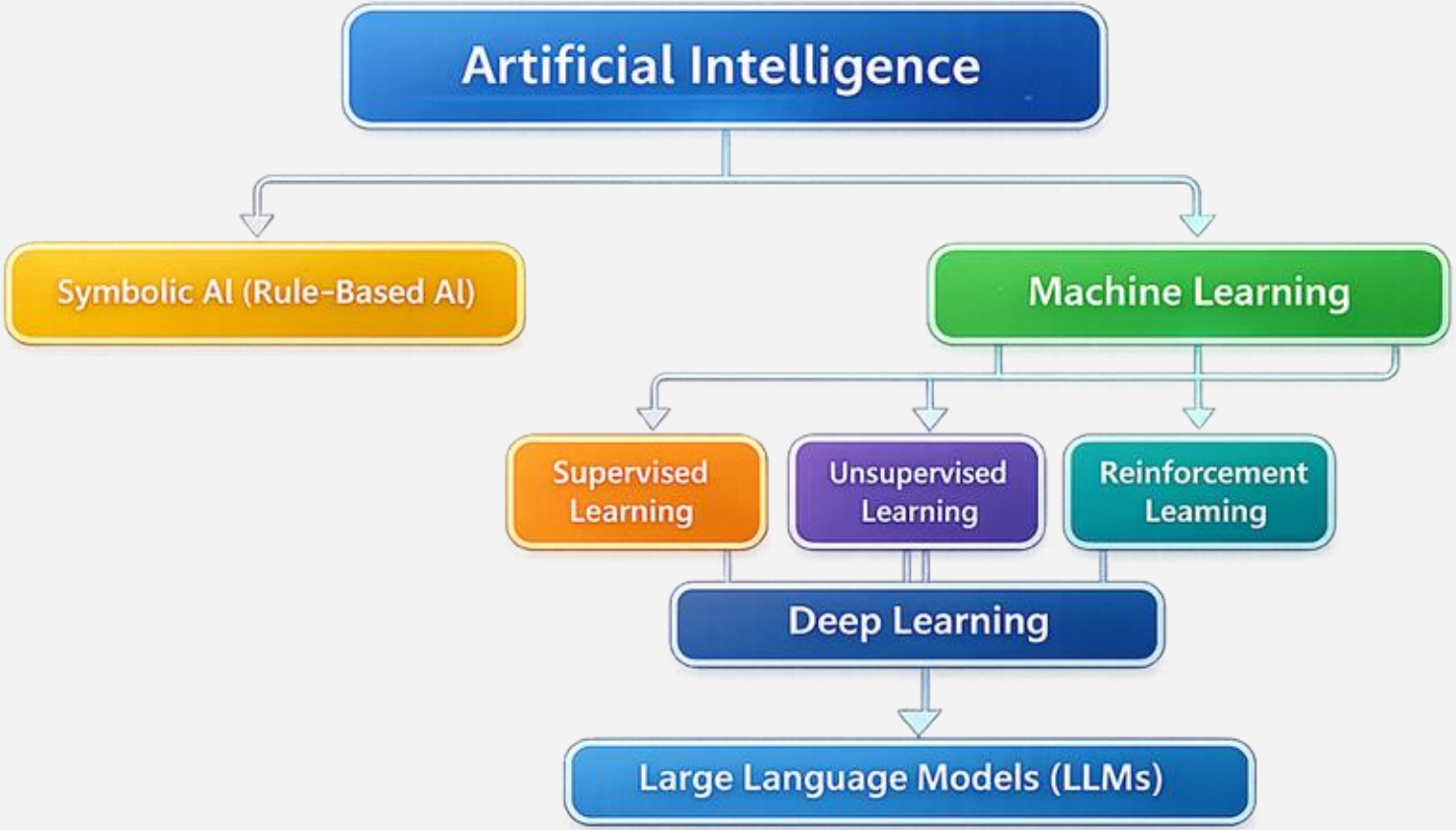
AI

Artificial Intelligence (AI)

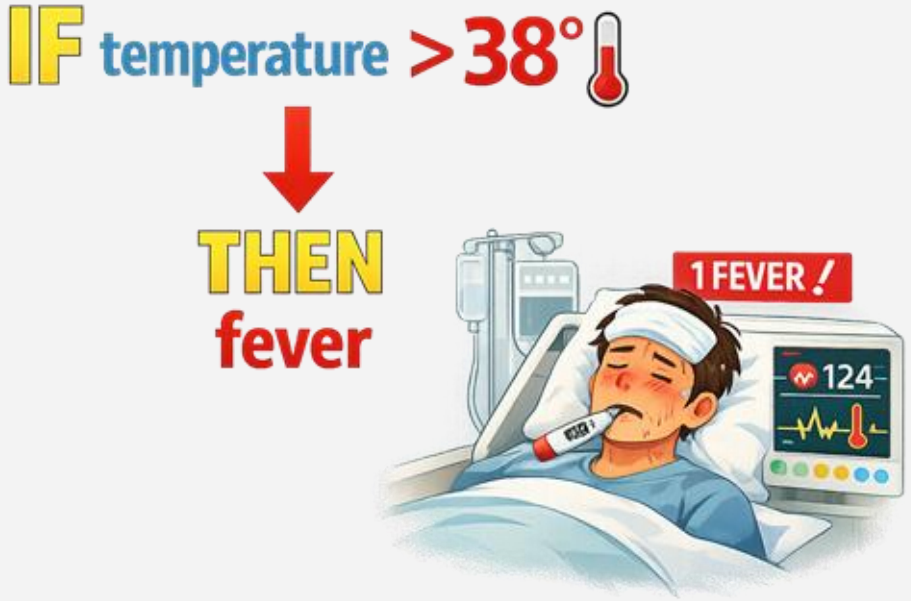
- Artificial Intelligence is the field of building computer systems that can perform tasks that normally require human intelligence.

- AI systems can:
  - Solve problems
  - Learn from data
  - Recognize images and sounds
  - Understand language
  - Make decisions




Artificial Intelligence
Symbolic AI (Rule-Based AI)
Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Leaming
Deep Learning
Large Language Models (LLMs)

Symbolic AI (Rule-Based AI)



Machine Learning (ML)

- Machine Learning is a **subset of AI** where systems **learn patterns from data instead of using fixed rules**.

- Main types of ML:
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning (RL)

Supervised Learning

- Learning from **labeled data**.

Predict house price

Classify emails as spam/not spam

Supervised Learning

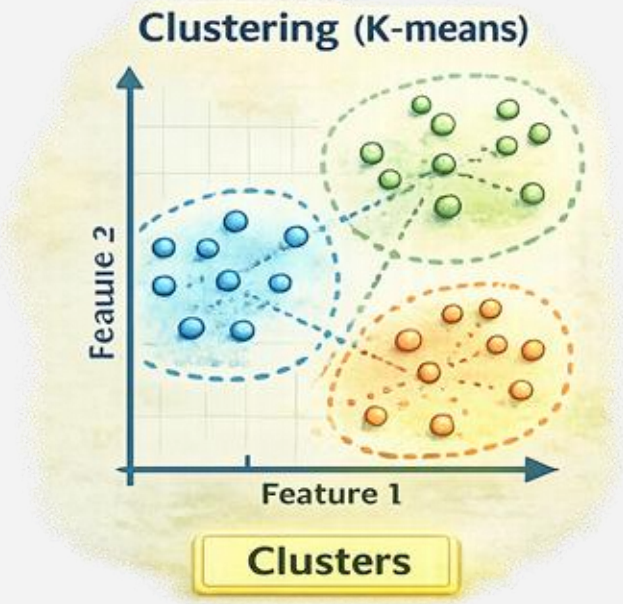
Unsupervised Learning

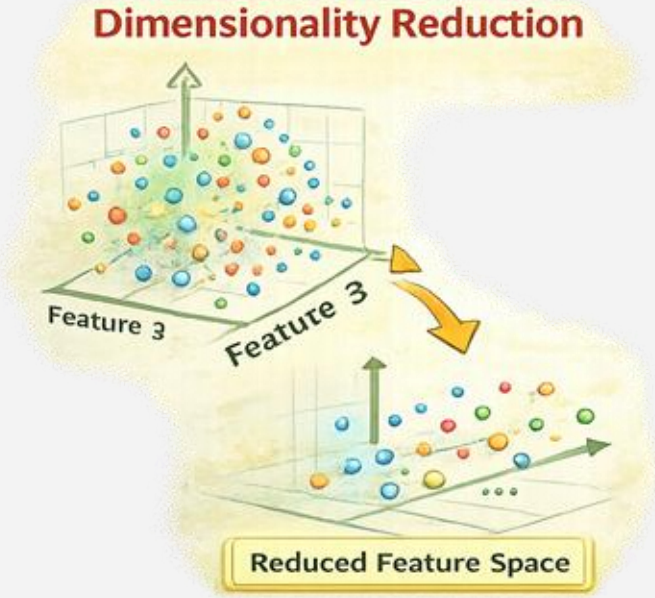
- Learning **patterns without labels**.

Customer segmentation

Anomaly detection

Unsupervised Learning





Reinforcement Learning (RL)

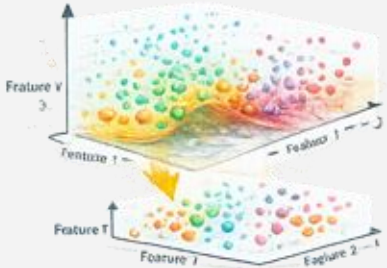
- Learning **through trial and error using rewards**.

Deep Learning

- Deep Learning is a **subset of Machine Learning** based on **neural networks with many layers**.

Image Recognition

Speech Recognition



Autonomous Vehicles

Natural Language Processing

Deep Learning

CNNs (images)

(images)

RNNs (sequences)

(sequences)

Transformers

(language models)

Large Language Models (LLMs)

- LLMs are a **specialized type of Deep Learning model** designed for **language understanding and generation.**

GPT (OpenAI) Gemini (Google) Claude (Anthropic) DeepSeek LLaMA (Meta)

Applications Based on LLMs

Domain	Modern Examples
General Purpose Assistants	ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google)
Coding & Engineering	GitHub Copilot, Cursor, Replit
AI Search Engines	Perplexity AI, Microsoft Copilot, Google AI Overviews
Vibe Coding	Loveable, Base44
Voice & Audio Generation	ElevenLabs, OpenAI Voice Mode, Suno (Music)
Image Generation	Nano Banana, Midjourney
Video Generation	Veo3 (google), Runway

The Pillars of Modern AI

Hype Cycle for AI Technologies