CMPE 252

AI & Data Engineering

Special Accommodations

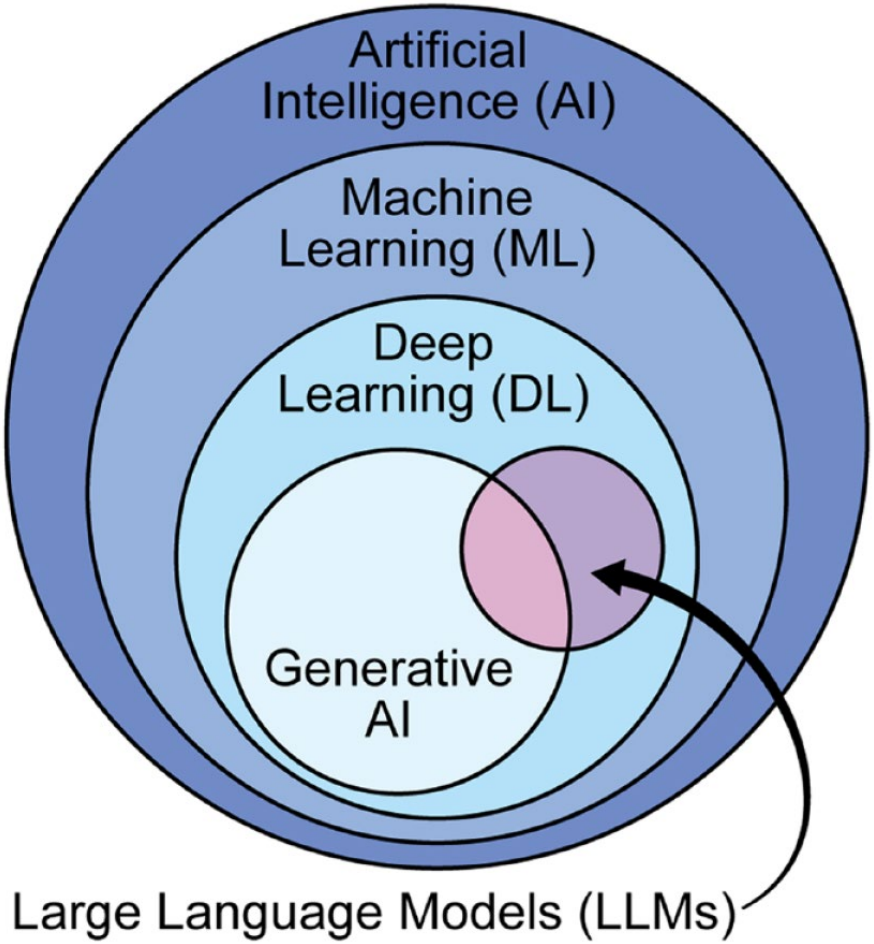
- Any student who feels that he or she may need an accommodation because of a disability (learning disability, attention deficit disorder, psychological, physical, etc.), please make an appointment to see me during office hours.

COURSE INTRODUCTION

- So, what's this course about?

- We will focus on the fundamental problems in AI
  - ❑planning, e.g., autonomous driving path planning
  - ❑graph search, e.g., shortest path
  - ❑programming, e.g., dynamic/linear/integer programming
  - ❑Learning, e.g., ML, DL, RL
  - ❑LLM, e.g., transformer, RAG
  - ❑data engineering
  - ❑etc.

Example


Artificial Intelligence (AI)
Machine Learning (ML)
Deep Learning (DL)
Generative AI
Large Language Models (LLMs)

Quick poll

1. Are you familiar with neural networks?
2. Are you familiar with reinforcement learning?
3. Are you familiar with LLM?

AI

Artificial intelligence is not just about learning methods like machine learning, deep learning, and reinforcement learning; it also involves using artificial intelligence ideas to address engineering challenges across different disciplines.

AI Definition

Turing test:

- The **Turing Test**, originally proposed by Alan Turing in 1950 as the "Imitation Game," is a method for determining whether a machine can exhibit **human-like intelligence (Artificial Intelligence)**.
- Rather than defining "thinking," Turing suggested that if a machine could convincingly mimic human conversational behavior, it should be considered intelligent.

Turing Test-How the Test Works

- The standard setup involves three participants who are physically separated and communicate only via text:
- **A Human Judge**: Asks questions to two hidden entities.
- **A Machine**: Attempts to convince the judge that it is the human.
- **A Human Respondent**: Provides a baseline for human-level conversation.
- If the judge cannot reliably tell which participant is the machine, the machine is said to have passed the test.

Key Benchmarks and Modern Status

- **2025 Milestone:** In June 2025, researchers at UC San Diego reported that **GPT-4.5** passed a rigorous Turing Test, being mistaken for a human **73% of the time**, compared to actual humans who were correctly identified only 67% of the time.

- **The "Turing Trap":** Critics argue that the test rewards **deception** and surface-level mimicry rather than genuine reasoning or consciousness. This has led some to focus on "The Turing Trap," where AI development prioritizes imitating humans over augmenting human capabilities.

- **Reverse Turing Test:** A common modern variation is the **CAPTCHA**, which is a "Reverse Turing Test" where a machine evaluates whether an interlocutor is a human or a bot.

Common Variations of Turing Test

- **Total Turing Test:** Adds requirements for the machine to demonstrate physical perception (vision) and object manipulation (robotics).
- **Visual Turing Test:** Specifically evaluates the machine's ability to identify details in images.
- **Lovelace Test:** Examines whether an AI can generate original, creative ideas that exceed its initial programming.
- **Marcus Test:** Evaluates if an AI can understand the underlying meaning, humor, or sarcasm in video content.

Current AI Models

- The **Turing Test** is no longer considered the "final frontier" for AI, as several frontier models passed it in 2025 by being mistaken for humans over **70% of the time**. Instead, the industry has shifted to **reasoning** and **agentic benchmarks** that test specialized human-level professional skills.
- The table in the next slide compares the top AI models as of now (January 2026) across key performance metrics:

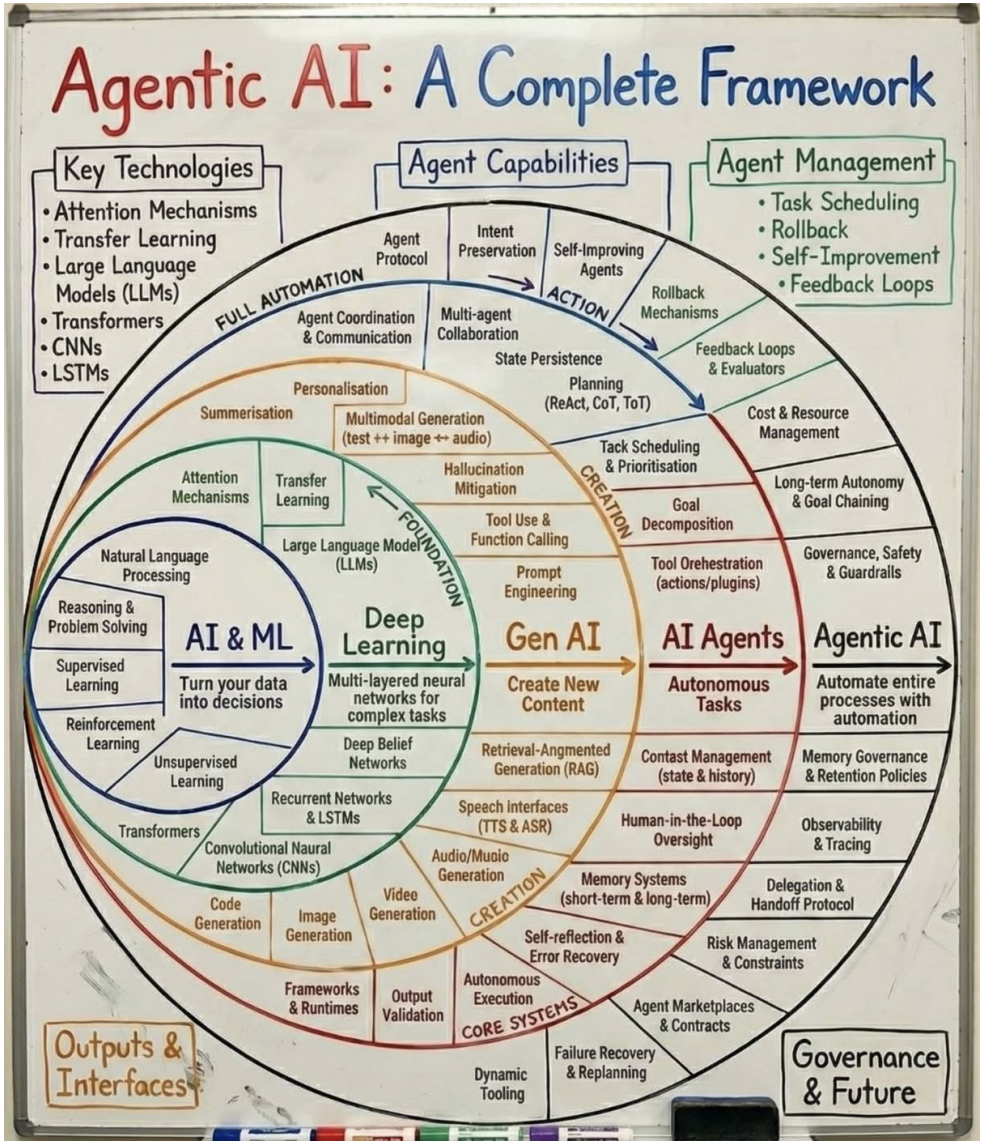
AI Model Comparison

Feature	GPT-5.2	Claude 4.5	Gemini 3 Pro
Primary Strength	Best Mathematical Reasoning & Problem Solving	Leader in Coding & Long-form Narrative	Top Multimodal & Ecosystem Integration
Turing-Era Benchmarks	Mistaken for human 73% of the time (GPT-4.5)	Dominates writing and strategic insight	#1 on LM Arena crowdsourced rankings
Logic/Math Score	100% on AIME 2025 high school math	86-87% on MMLU reasoning	100% on AIME 2025
Coding Success	80% on SWE-bench (Agentic Coding)	80.9% on SWE-bench (Industry leading)	76.2% on SWE-bench
Deep Reasoning	52.9% on ARC-AGI-2 (Abstract problems)	50%+ success on 5-hour software tasks	Excels at complex STEM questions

Key Trends for 2026

- **The "Trust" Benchmark:** Standard tests like MMLU are being replaced by **"Competence Trust,"** a measure of how accurately AI completes multi-step tasks without hallucinations.
- **On-Device Reasoning:** New models like **LFM2.5-Thinking** (January 2026) can now perform advanced reasoning locally on a phone, fitting within 900MB of memory.
- **Professional Failure Points:** Despite passing basic imitation tests, even the best models currently fail **Apex-Agents**, a 2026 benchmark that tests a model's ability to navigate professional tools like Slack and Google Drive simultaneously.
- **Passing Humans:** In many reasoning benchmarks, current models now outperform actual humans; for instance, real humans identified each other correctly only 67% of the time in the same tests where AI scored over 70% human-likeness.

AI Landscape



AI Application Example

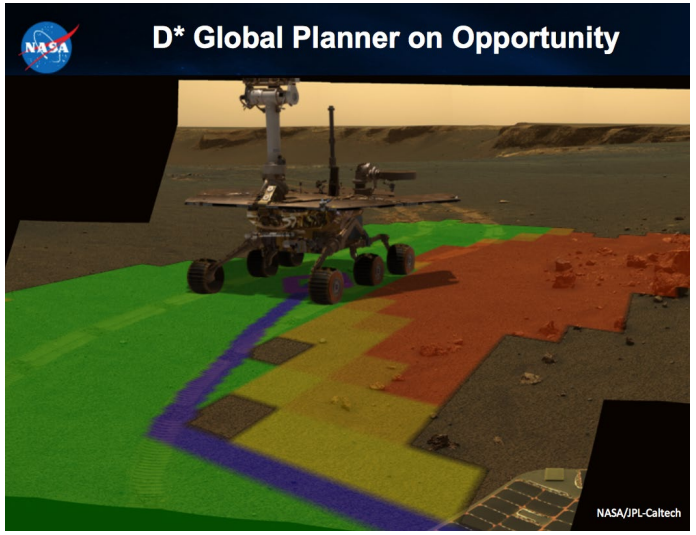
- How should the robot go from point A to point B?

- We will start simple:
  - Point robot that can move in any direction
  - Known environment with stationary obstacles
  - Perfect sensing
  - Perfect control

Example

- How should the robot go from point A to point B?

- We will start simple:
  - Point robot that can move in any direction
  - Unknown environments and/or moving obstacles
  - Perfect sensing
  - Perfect control

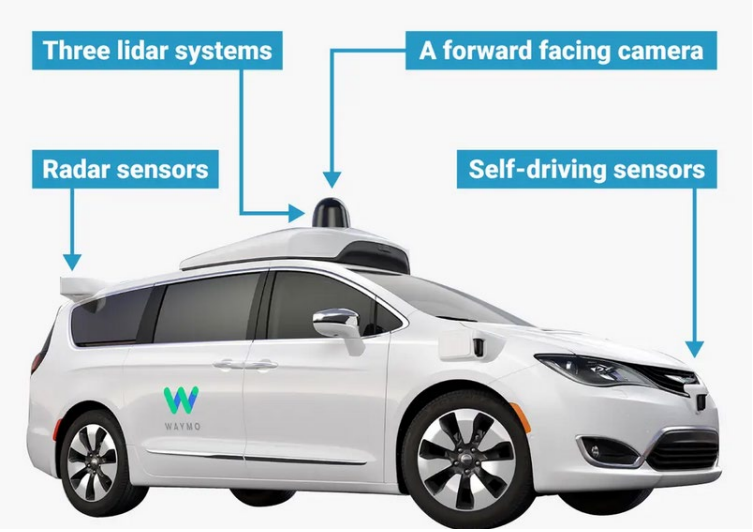


- Challenges?

Example: path planning – autonomous driving

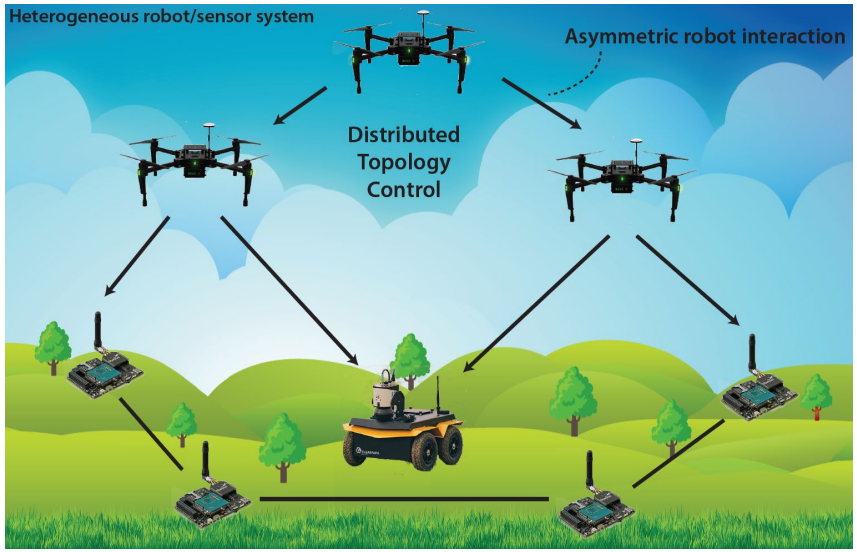
- How should the robot go from point A to point B?

- We will start simple:
  - Point robot that can move in any direction
  - Complex, high dimensional environments
  - Perfect sensing
  - Perfect control



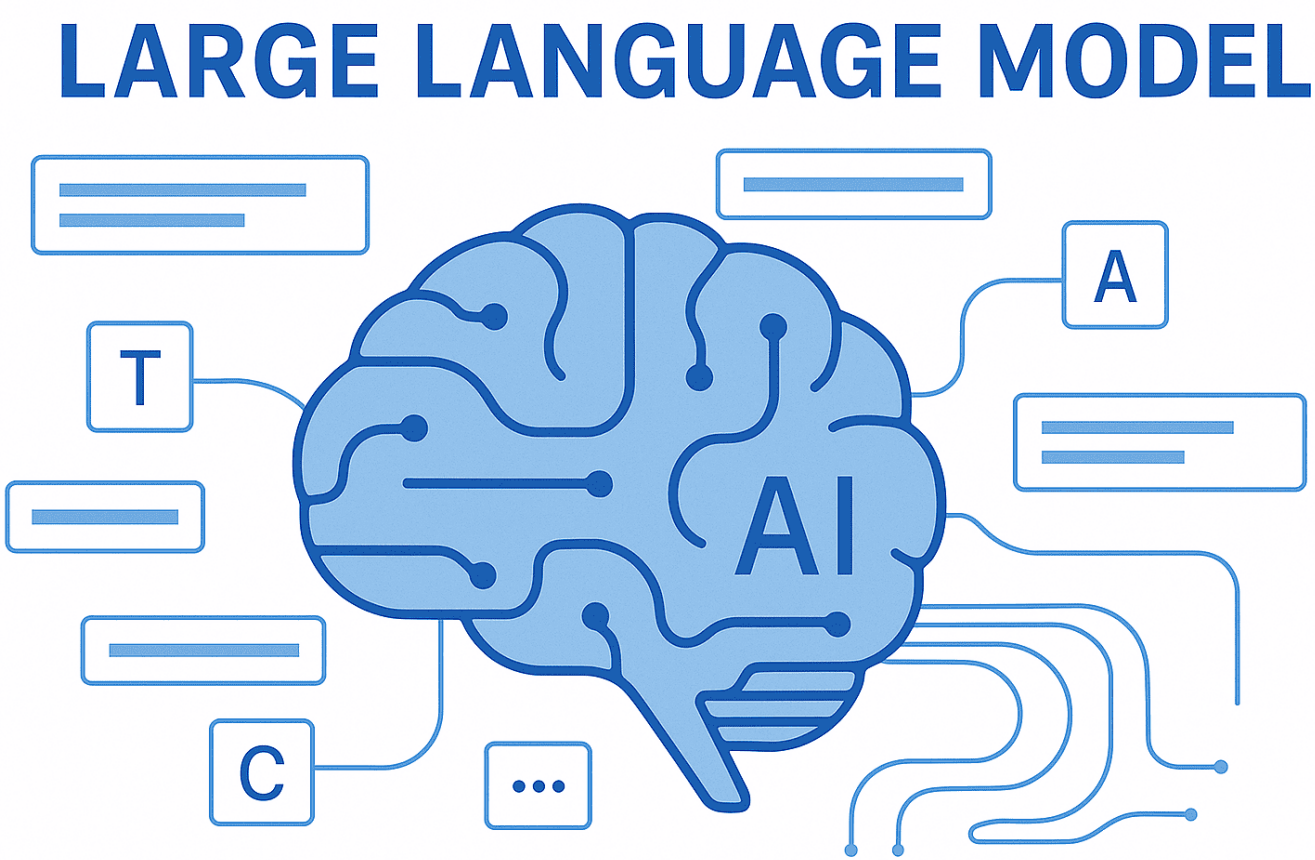
- Challenges?

Example: heterogeneous multi-robot systems



What are the challenges?

Example



Prerequisites

- Comfortable with data structures
  - graph, tree, etc.
- Basics of probability
  - typical distributions, pdf, mean and covariance of random variables
- Comfortable programming in at least one of MATLAB, Python, C/C++
  - will be implementing algorithms discussed in class
  - there will be NO coding support

CLASS FORMAT

- Homework:
- Paper presentation:
  - select from a list, or propose a new one (w/ requirements)
- Midterm:
  - In-class. closed books
- Final Project:
  - Proposal (w/ short presentation) + presentations + report

** The format is subject to change depending of enrollment numbers*

GRADING *

- Participation/attendance: 5%
- Homework: 20%
  - 3~ 5 assignments
- Midterm: 30%
- Paper implementation: 15%
- Final Project: 30%
  - Proposal (w/ quick presentation) + presentation + implementation + report

** subject to change depending of enrollment numbers*

PAPER SELECTION

- A google document will be used for paper selection.
- Selections will be first-come, first-served.

ACCESSING PAPERS

- Paper list will be posted on class website (more on this later).
- All papers should be accessible through standard sources, e.g.:
  - Google Scholar
  - IEEE Xplore
  - arXiv
- For details on accessing research publications go to library.sjsu.edu.

PAPER IMPLEMENTATION

- Presentations should convince the audience that you wrote the paper.
- We are not interested in a copy of the paper!
- Tell us:
  - why the paper is important
  - how it innovated
  - how you think it could be applied or extended.
- Again, presentations must be ~15 minutes in length.
- Everyone in class will be expected to have read the abstract and introduction of the paper.
- Participation in the discussion time after each presentation will be expected throughout the year.

PROJECT

- proposal: 1-2 pages, IEEE conference format, word or latex, template will be provided
- Please describe what you want to do, what the challenges are, and why they are important.
- Important: plan, deliverables.

PROJECT

- The project report must be submitted as a paper of 6 pages (standard IEEE conference format) describing the project outcome.
  - It is fine if you use some part of your proposal.
- There project requirements will be given.
- The project should ideally be the core of a potential conference paper submission. Students can come to me to discuss ideas and to brainstorm.

PROJECT

- everyone/group will give short/quick proposal presentation.
- This presentation is used as a guideline to pace your final project.
- Presentations should focus on what you will do
- Presentations can be 5-10 minutes.
- Participation in the discussion time after each presentation will be expected throughout the semester.

PROJECT IMPLEMENTATION

- Your final project must be accompanied by an implementation.
- Use the framework of your choice.
- Implementation will allow for visuals in the presentation and a deeper understanding of the material.
- Implementations & report need to be submitted.

LOGISTICS

- The course material will be on Canvas
- Make sure to check canvas & the online schedule regularly for updates
- Ensure that you are signed up for notifications