Introduction

Course AIAA 4220

Week 1 - Lecture 1

Changhao Chen

Assistant Professor

HKUST (GZ)

Instructor

Changhao Chen

- Assistant professor,
- PhD, University of Oxford (2020).

Email: changhaochen@hkust-gz.edu.cn

Experience & Education:

- Assistant Professor, AI and INTR Thrust, The Hong Kong University of Science and Technology (Guangzhou), Jan 2025 - Now
- Lecturer, National University of Defense Technology (NUDT), Dec 2020 – Dec 2024
- Postdoctoral Researcher, University of Oxford, Feb 2020 – Oct 2020
- Ph.D. in Computer Science, University of Oxford, Oct 2016 - Oct 2020
- Master in Control Science and Technology, National University of Defense Technology (NUDT), Sept 2014 – Sept 2016

Instructor

Changhao Chen

- Assistant professor,
- PhD, University of Oxford (2020).

Email: changhaochen@hkust-gz.edu.cn

Research Interests:

3D Scene Perception
System State Estimation
Embodied AI
Efficient Neural Computing

Awards:

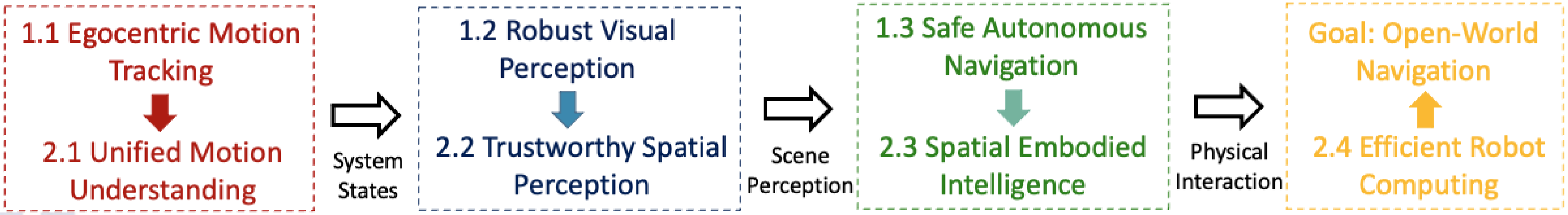
- World's Top 2% Scientists (2024)
- Young Elite Scientist Sponsorship Program, CAST (2022)
- Robotics: Science and Systems Pioneers (2020)

Research Outcomes:

- PI for NSFC General Program and Youth Program
- Published over 50 high-quality papers in the fields of AI, intelligent transportation, and robotics, including top journals such as TITS, TNNLS, TMC, and TIP, as well as top-tier conferences including CVPR, ICCV, AAAI, ECCV, ICRA, IROS, with over 3,200 Google Scholar citations.
- Granted 14 patents, including international PCT patents, and patents from China, the US, Europe, and Australia, including one successfully commercialized in the UK.

Our Research

- HKUST (GZ) PEAK-Lab is a research group dedicated to embodied AI and autonomous systems.
- Traditional robotic algorithms often depend on meticulously crafted physical and geometric models, which may struggle to the ever-changing complexities of the open world.
- Our mission is to bring AI into the physical open world by enabling machines to understand general motion, perceive 3D scenes, and actively navigate and interact with their surroundings.



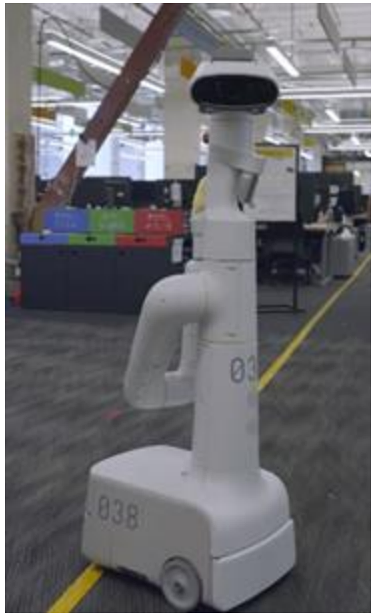
Homepage: changhao-chen.github.io/

PEAK Lab

Perception, Embodiment, Autonomy, Kinematics

What is Embodied AI?

Boston Dynamics



Everyday Robot

Unitree Robotics

Mobile manipulation

Various robots:

Two-legged, wheeled, four-legged…

What is Embodied AI?

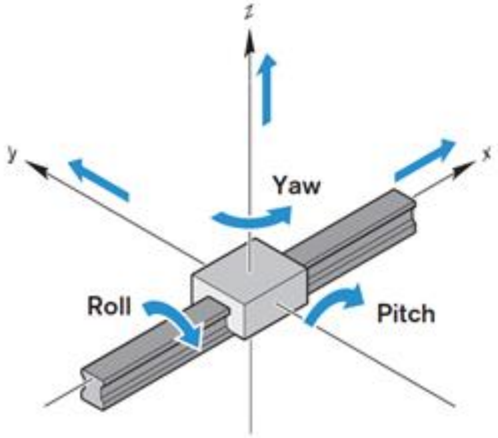
- Embodied AI is a paradigm in AI where an agent (software or hardware) learns to perform tasks by sensing, interpreting, and acting within a physical or simulated environment.

- Unlike traditional AI models that process static data (like images or text), embodied agents have a "body" (virtual or real) that allows them to interact with and change their surroundings to achieve a goal.

- It‘s about moving from “seeing ” "thinking" to "doing."

When Machine Learning is not Needed

- Industrial Robot
  - Populated from 1970s
  - Usually a single arm with no mobility
  - Move an object with up to 6 DoF
  - Programmed to do repetitive actions
  - Repetitive, non-changing processes



Define points P1–P5:

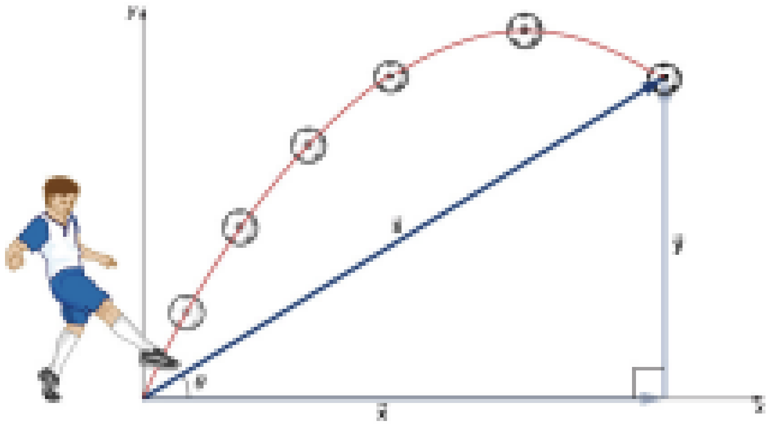
1. Safely above workpiece (defined as P1)
2. 10 cm Above bin A (defined as P2)
3. At position to take part from bin A (defined as P3)
4. 10 cm Above bin B (defined as P4)
5. At position to take part from bin B. (defined as P5)

Define program:

1. Move to P1
2. Move to P2
3. Move to P3
4. Close gripper
5. Move to P2
6. Move to P4
7. Move to P5
8. Open gripper
9. Move to P4
10. Move to P1 and finish

Physical Model based Solutions

- Physical Modelling
  - Kinematics and inverse kinematics of different actuators
  - Meaning: how robots (parts) move to a desire position
  - We don’t have to “reinvent the wheel”. We have good models of physical processes



When Machine Learning is Needed

- Robots in **unstructured and dynamic environments**

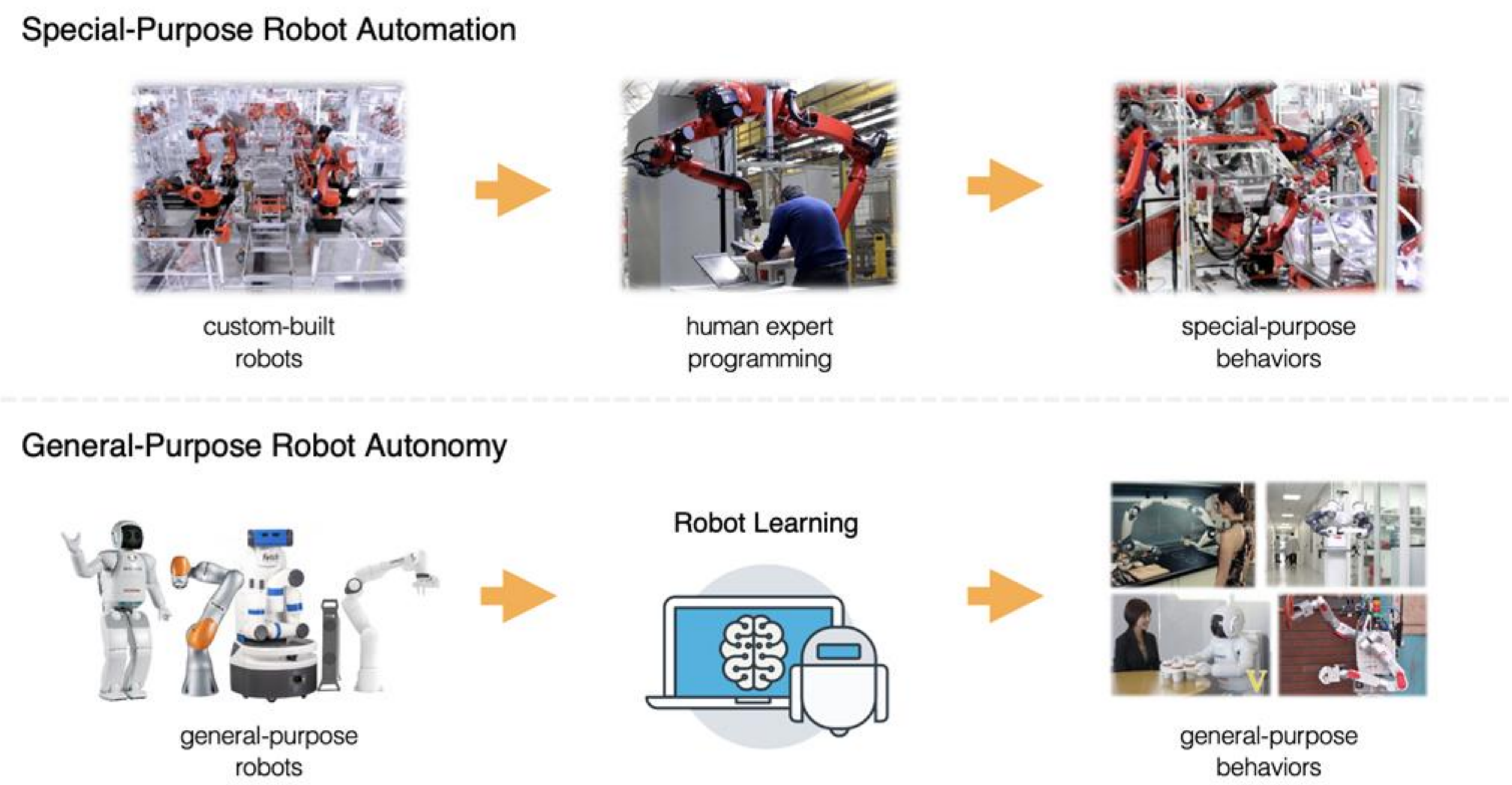
Environment uncertainty

Scene dynamics

Unknown Object

Complex tasks

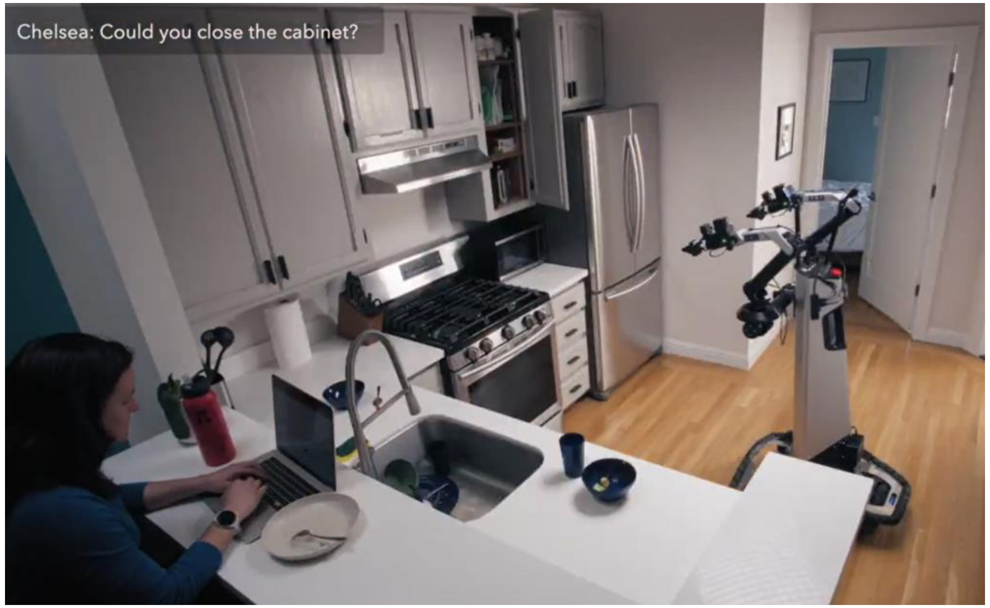
Embodied AI vs. Automation



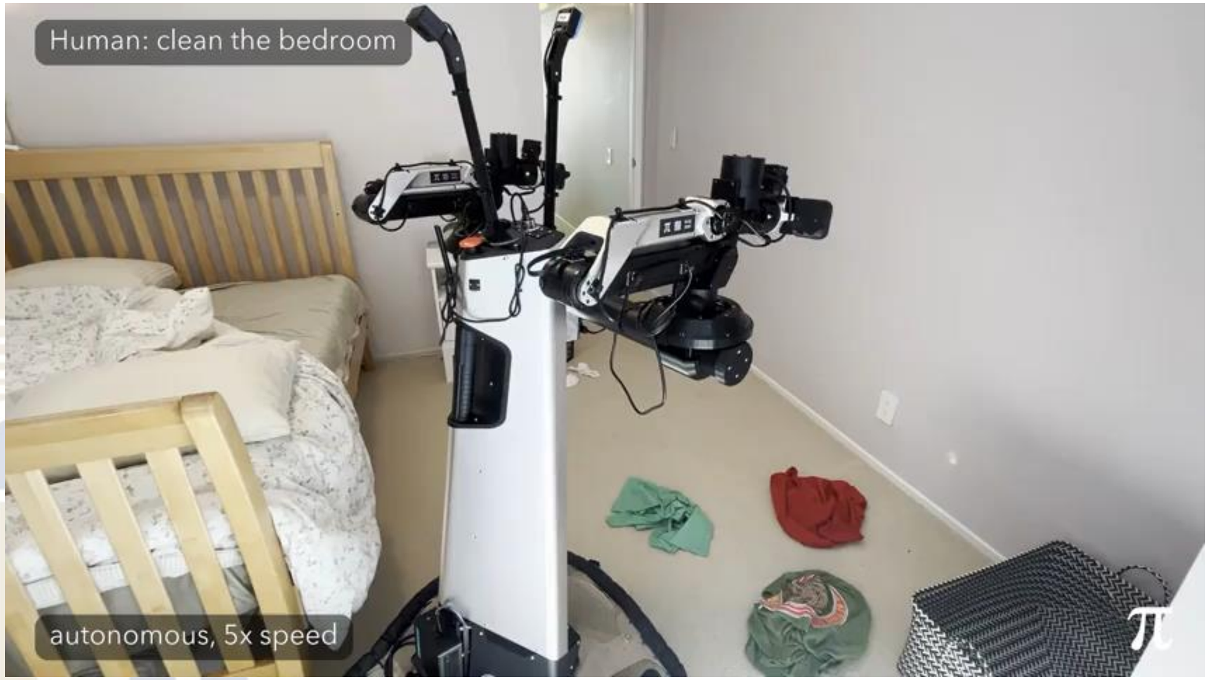
Recent Advancements

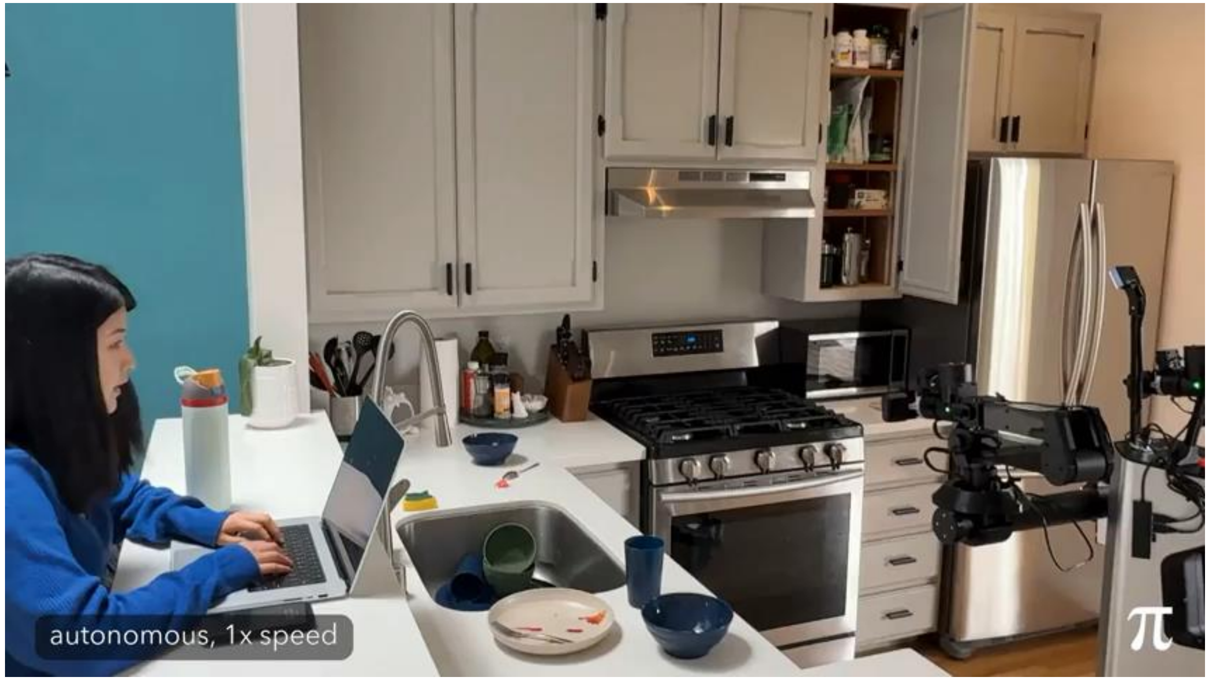
- Warehouse Robot
  - Boston Dynamics 2022 new product
  - Unloading boxes from trucks to warehouses
  - Adapt to many kinds of boxes

- Service Robot
  - Pi 0.5 model from Physical Intelligence 2025



Recent Advancements

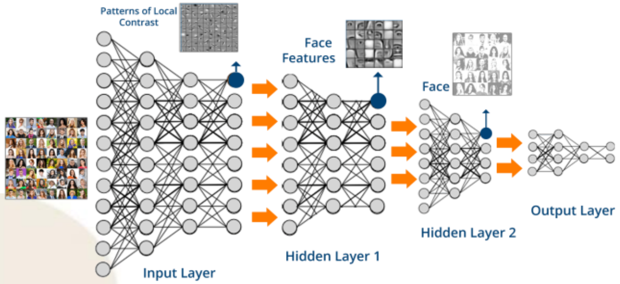




Embodied AI Applications

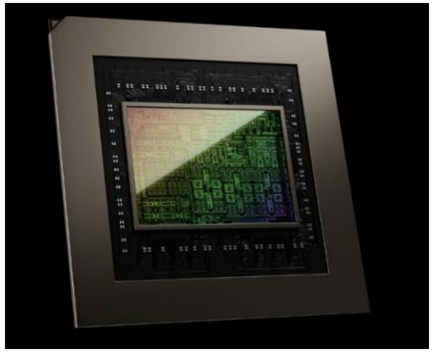
- Cleaning robots
- Service robots
- Self-driving vehicles
- Delivery robots

Now is a good time to study and work on Embodied AI



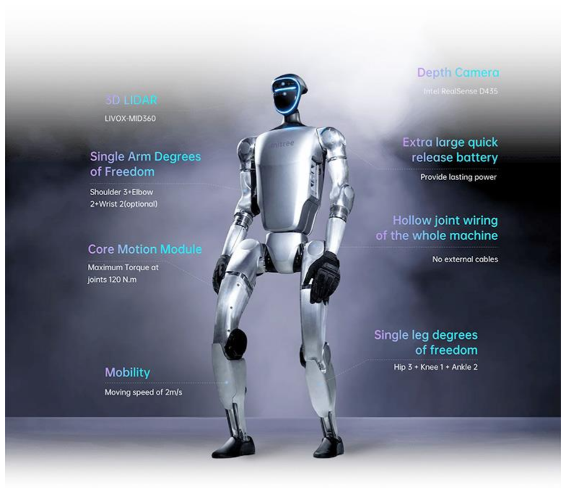
Artificial Intelligence

Recent breakthroughs in machine learning, language and computer vision: CLIP, ChatGPT. Expect a good Large Vision-Language Model in recent years



Computing Power

Your smartphone is millions of times more powerful than all of NASA's combined computing in 1969.



Robot Hardware

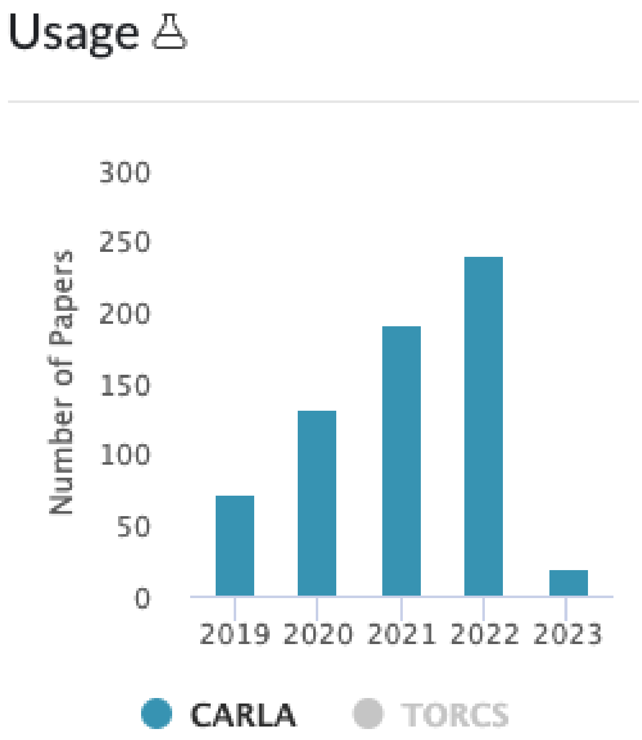
More reliable and affordable robot hardware. Boston Dynamics' spot costs $75,000
Unitree has $10,000 ones

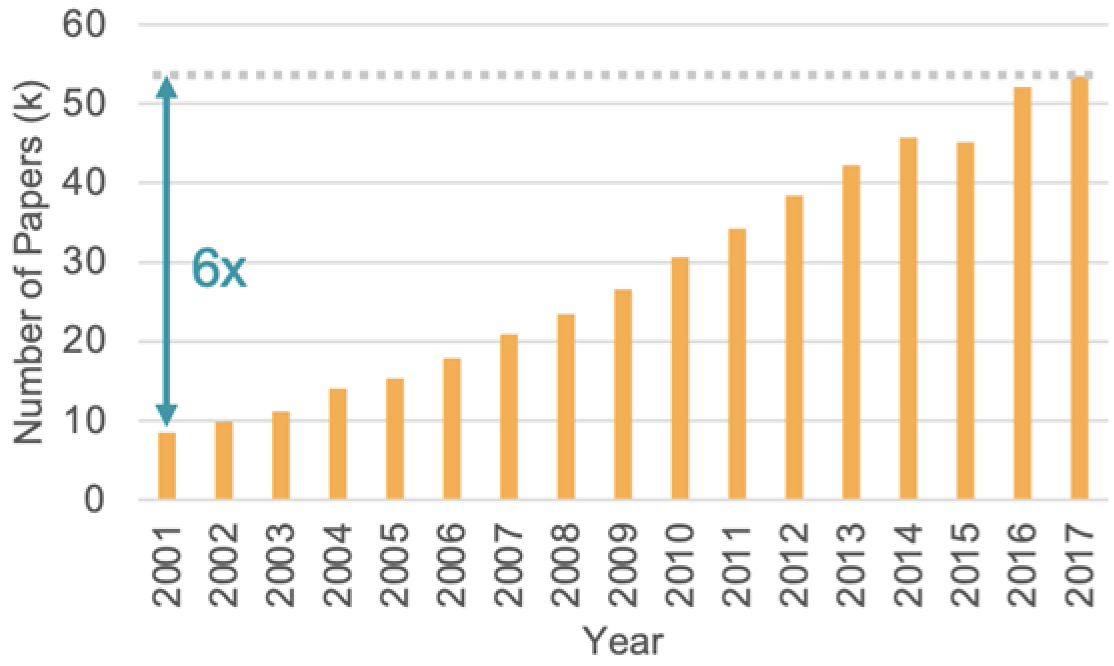
3D Simulators

More open-source, easy-to-use 3D simulators.
*CARLA
*NVIDIA Isaac Sim

Now is a good time to study and work on Embodied AI

- The research community is rapidly growing
    - Publications at top conferences are on the rise!
        - ICRA, CoRL, RSS, IROS, etc.





Embodied AI is challenging!

From DARPA Robotics Challenge 2015

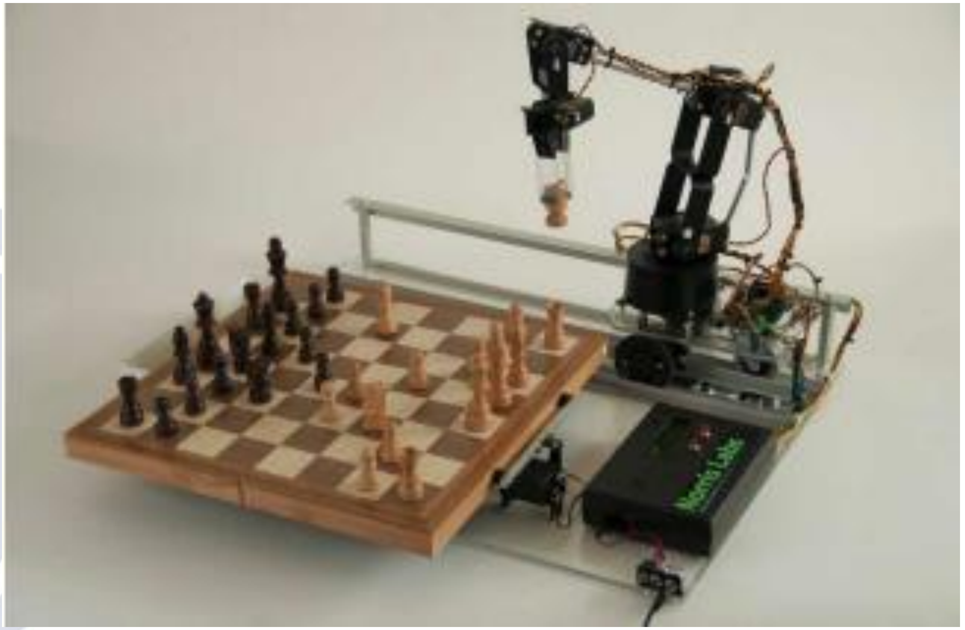
A 3-stage competition to develop robot assistants. Tasks involved:

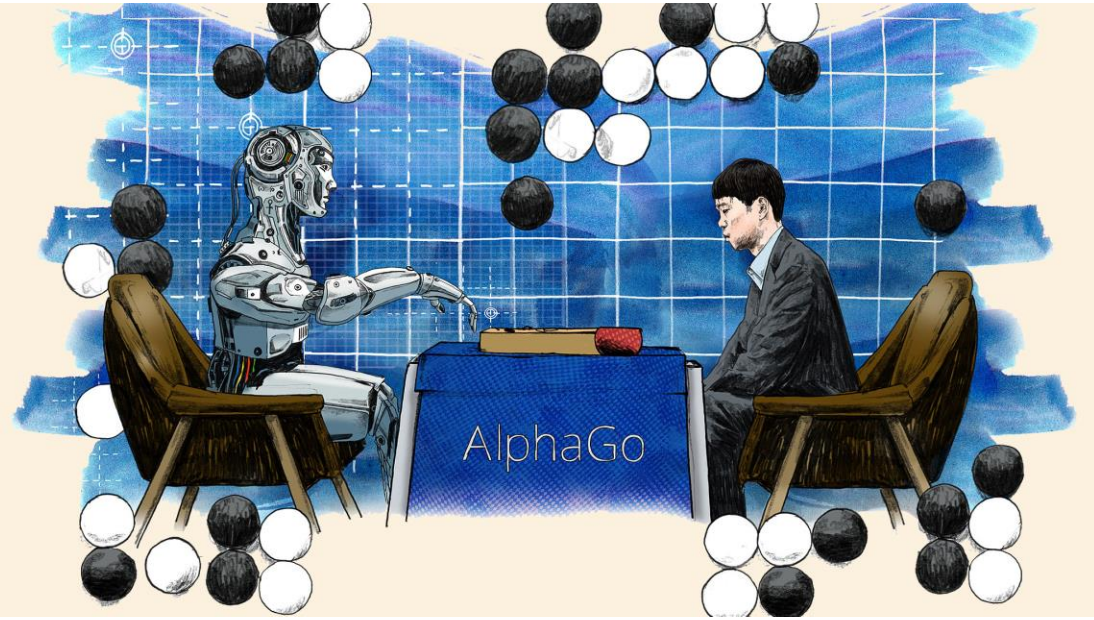
- Robot drives through an obstacle course to destination and exit car
- Enters a building through door
- Turns a valve
- Cuts a hole in a wall using a tool
- Climb stairs to finish

Only 3 teams successfully finished all the tasks

Embodied AI is challenging!

- The Moravec's paradox (1988)
  - “it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility”





Key Components in Embodied AI

Perception

Seeing and understanding
of surroundings

Decision Making

Motion planning and control
of robot behaviors

Robot Systems

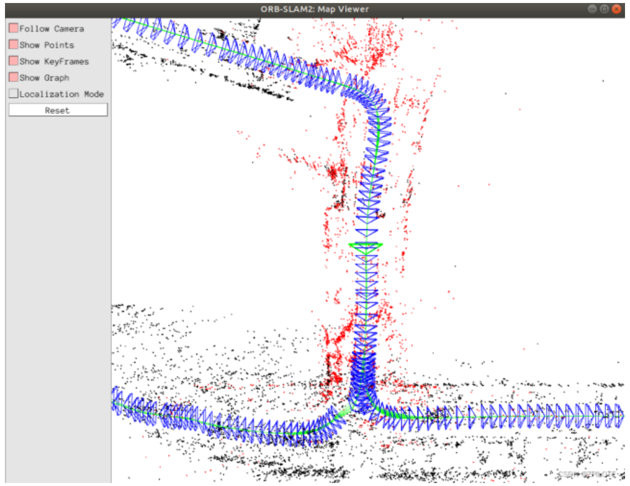
Robot embodiment.
3D simulations.
Sim2real

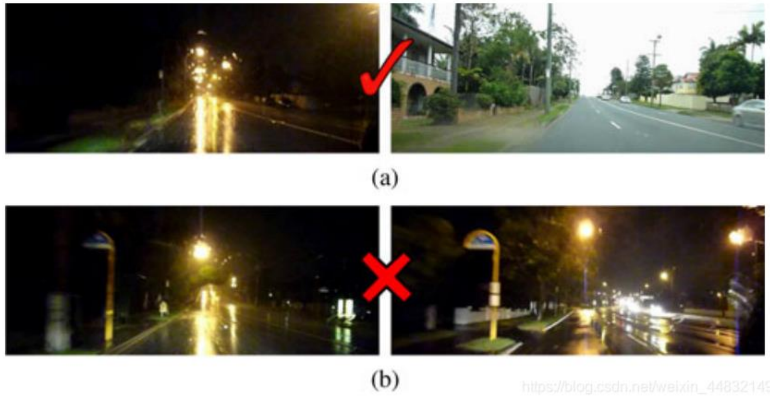
Embodied AI - Perception

Object Detection and Tracking

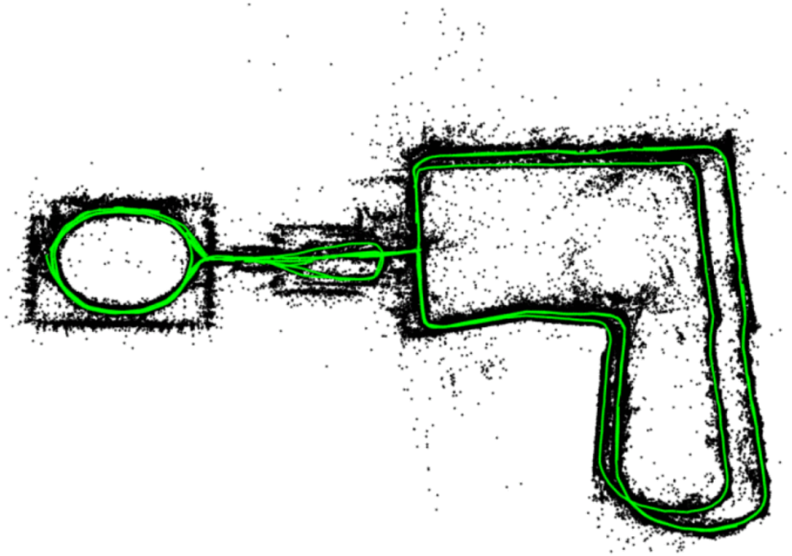
Semantic Understanding

Embodied AI - Perception





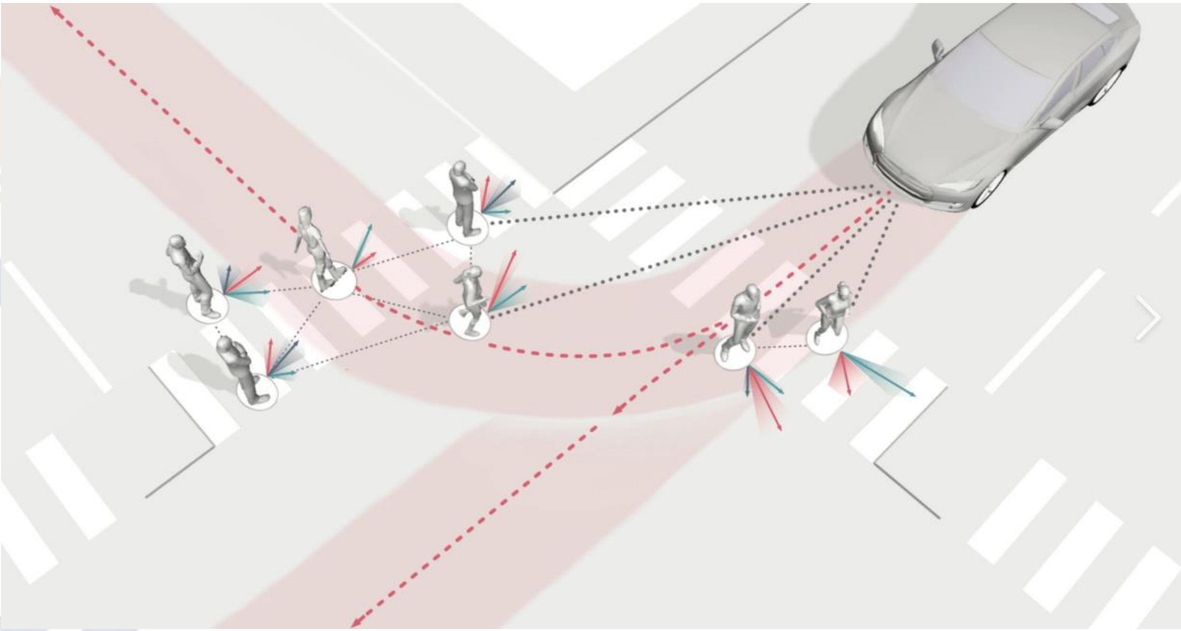
Localization- Place Recognition



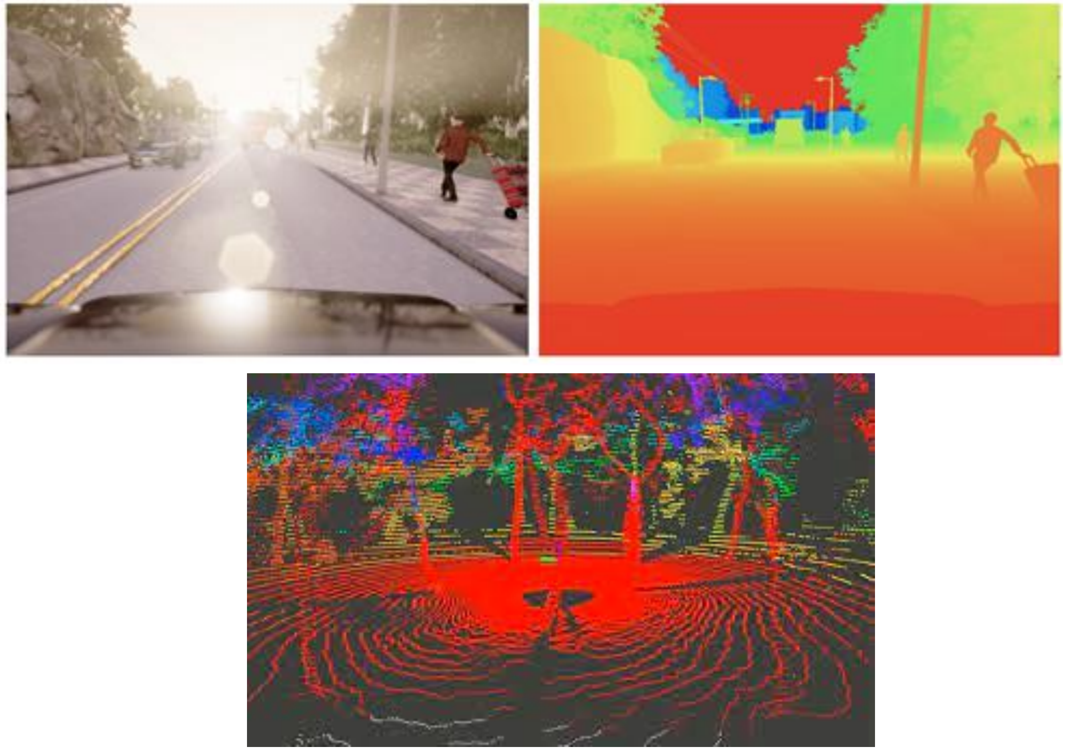
Mapping

Embodied AI - Perception

- Trajectory Prediction
  - Predict the future trajectory/locations of vehicles and pedestrians

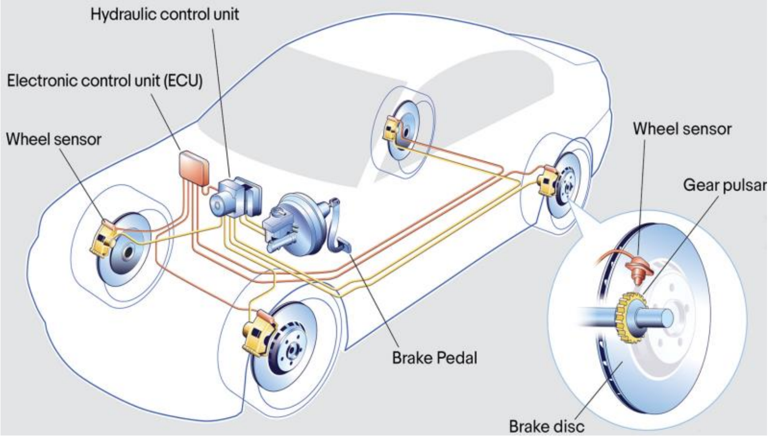


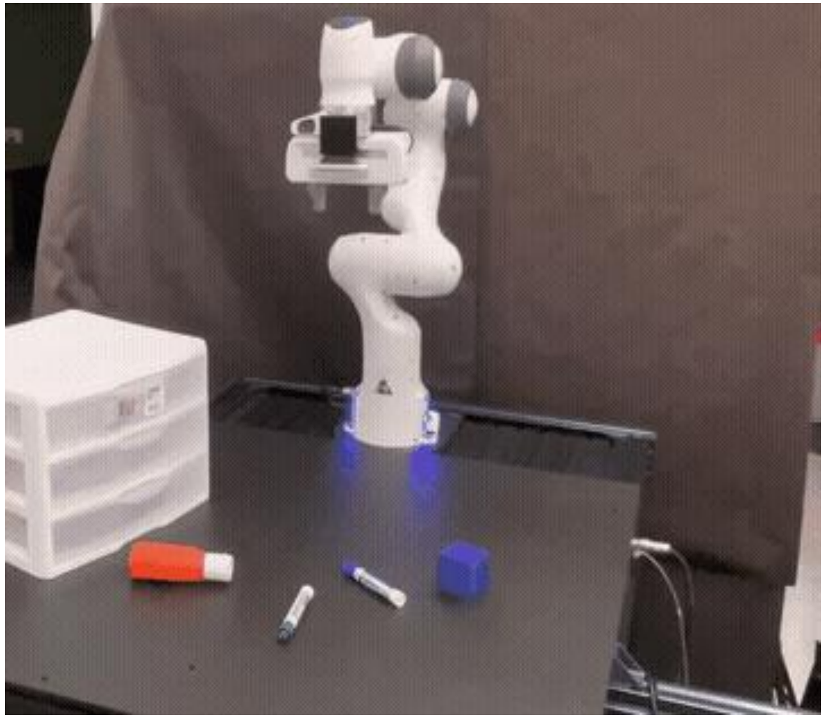
- Sensor Fusion
  - Combine information from RGB, Depth, and point clouds (LIDAR)



Embodied AI - Decision Making

- Navigation
- Manipulation
- Motion Planning and Control

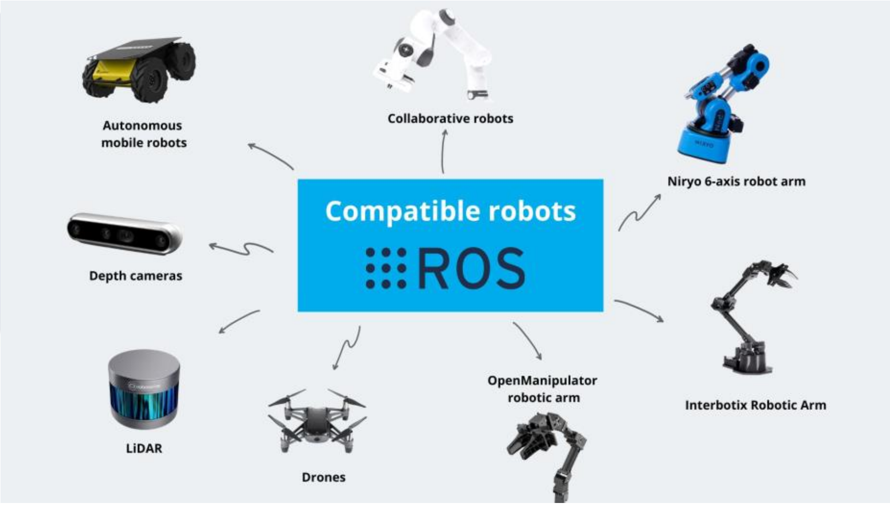


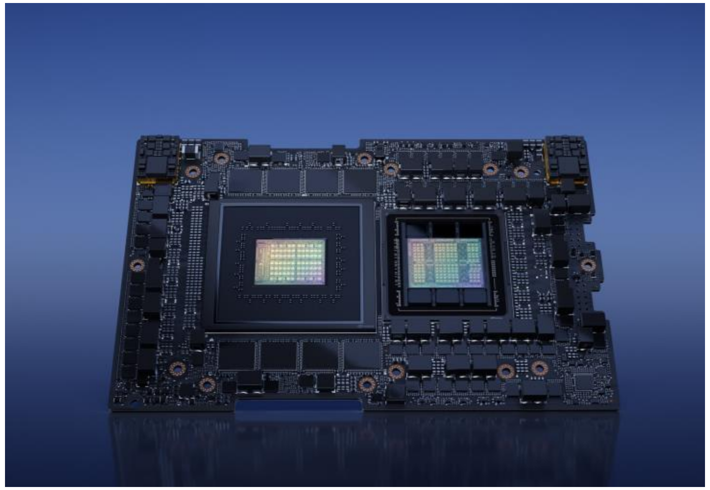


*RVT: Robotic View Transformer for 3D Object Manipulation. CoRL 2023

Embodied AI - Robot Systems

- Systems and Hardware

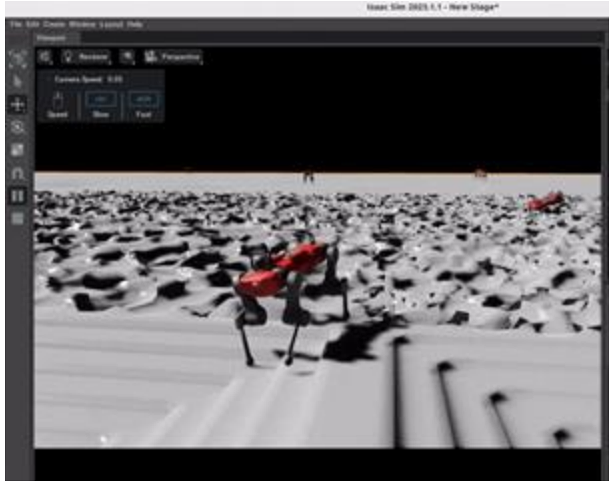


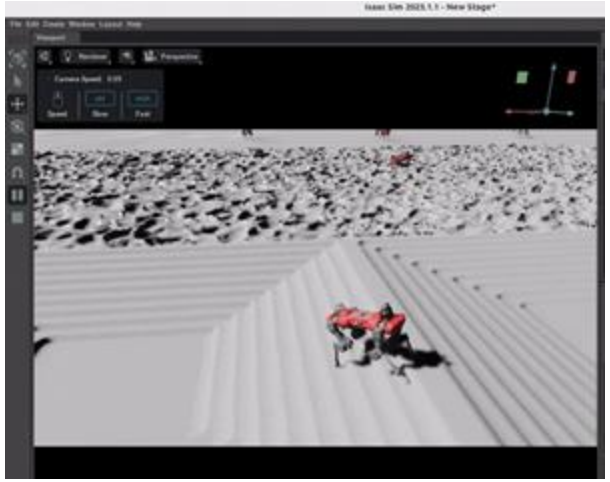


Embodied AI - Robot Systems

- 3D Simulation

* CARLA simulator (based on UE4): https://carla.org/ https://leaderboard.carla.org/





* Nvidia Isaac Sim: https://developer.nvidia.com/isaac-sim

Course Roadmap

- Machine Perception
  - Computer Vision
    - Visual Recognition
    - Object Tracking
- Robot Systems
  - 3D Simulators
  - PyTorch
  - Robot Operating System (ROS)
- Decision Making
  - Embodied Navigation

    Localization and Mapping, Reinforcement Learning, etc.
  - Embodied Interaction and Manipulation

    VLA, etc.

Week	Lecture	Date	Topics	
1	1	Wednesday, Sep 3	Introduction	Introduction
1	2	Wednesday, Sep 3	Embodied AI Overview	
2	3	Wednesday, Sep 10	Intro to Computer Vision	Robotic Perception
2	4	Wednesday, Sep 10	Camera Models & Calibration	
3	5	Wednesday, Sep 17	Visual Recognition I	
3	6	Wednesday, Sep 17	Visual Recognition II	
4	7	Wednesday, Sep 24	Object Detection	
4	8	Wednesday, Sep 24	Object Tracking	
5	9	Saturday, Oct 11	Pytorch Basics	Robotic Systems
5	10	Saturday, Oct 11	Intro to ROS and 3D Simulator	
6	11	Wednesday, Oct 15	Project 1 Presentation	Embodied Navigation
6	12	Wednesday, Oct 15		
7	13	Wednesday, Oct 22	Localization and Mapping	
7	14	Wednesday, Oct 22		
8	15	Wednesday, Oct 29	Planning and Control	
8	16	Wednesday, Oct 29		
9	17	Wednesday, Nov 5	Reinforcement Learning	
9	18	Wednesday, Nov 5		
10	19	Wednesday, Nov 12	End-to-End Navigation	
10	20	Wednesday, Nov 12		
11	21	Wednesday, Nov 19	LLM & RLHF	Embodied Interaction and Manipulation
11	22	Wednesday, Nov 19		
12	23	Wednesday, Nov 26	Visual-Language Action Model	
12	24	Wednesday, Nov 26		
13	25	Wednesday, Dec 3	Project 2 and 3 Presentation	
13	26	Wednesday, Dec 3		

Course Lab

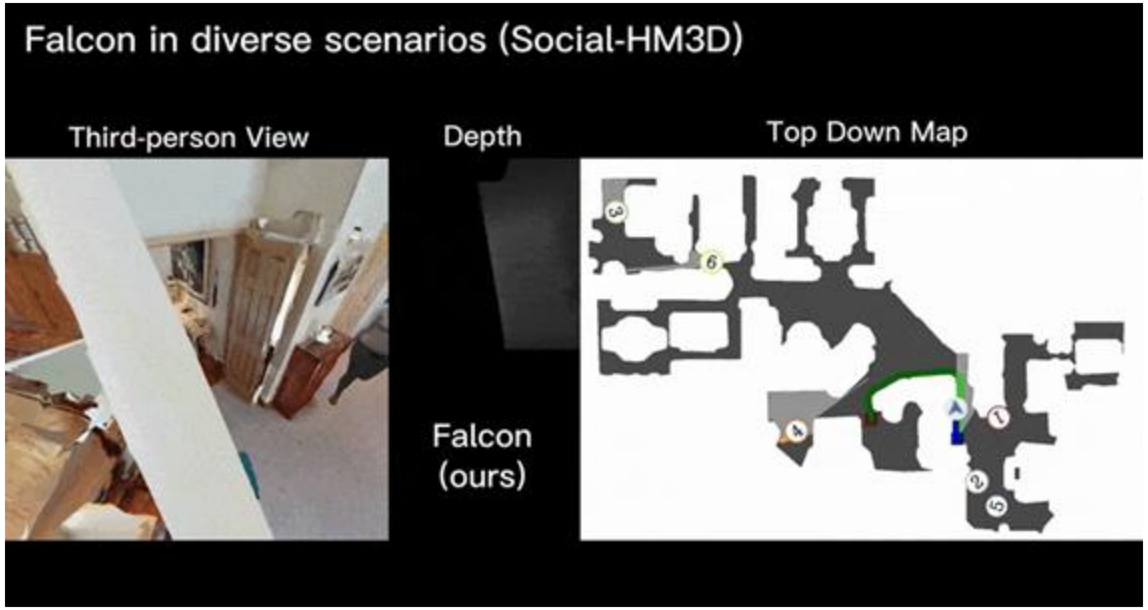
- Course evaluation overview
  - Project 1: Paper Presentation
  - Project 2: Egocentric Object Detection
  - Project 3: Robot Navigation

- Requirements
  - Python programming
  - Basic calculus, linear algebra, probability theory
  - You should really be using Ubuntu (not Windows)

- Hardware requirement
  - For project 3, we need a machine with at least 8GB GPU
  - So at least 3060 TI / 4060 laptop (don’t buy 50x0 series GPU)
  - There some free resources, but I just don’t think it is not enough/convenient to run Sim
    - Google Colab, Kaggle free GPU
    - 阿里云：https://university.aliyun.com/ free ¥300 RMB credit

Course Objectives

- Describe key principles, concepts, models and techniques in Embodied AI
  - Understand technical challenges and recent research progress
- Implement algorithms introduced in class to solve small real-world problems
  - Familiarity with Embodied AI systems like ROS and 3D simulators
- Develop novel methods for Embodied AI applications
  - Navigation
  - Manipulation



Computation Resources from Aliyun

- Aliyun has provided our university with free or discounted computation resources
  - https://university.aliyun.com/
  - 300 RMB voucher
  - 70% off for orders up to 5000 RMB/year
  - You can directly apply with your hkust-gz email

Grading

- This course utilizes an **absolute grading** system as follows
  - A:[100,85] B:(85,70] C (70,55] D:(55.40] F:(40.0]
  - Subgrades (e.g, A-, B+, etc,) will be assigned foreveny 5 points within these ranges.
  - If the grades are too low, we may **curve up** the whole class

- In-class Quizzes (25%)
- Project 1: Paper Presentation (20%)
- Project 2: Machine Perception Homework (20%)
- Project 3: Navigation Homework (30%)
- Attendance (5%)

Grading - In-class Quizzes

- In-class Quizzes (25%) + Attendance (5%) = 30%
  - Designed to complement the lectures and making sure you understand the key knowledge about the course
  - 5 In-Class Tests (5% each)
  - **Usually scheduled at the end of the class (around 4:10 pm.), due in-class**
  - All quizzes will be single or multiple choice questions. You will have 10 minutes (tentative) to finish.

Grading - Paper Presentation

- Project 1: Paper Presentation (20%)
  - You will select 1 paper from our list of Embodied AI research to present. This should show that you understand the paper, and you could present it to students who do not know the paper well. **You have 10 minutes to present and 5 minutes to handle at least one question (from TA/peers/teacher)**
    - Will be released on Sep 10
    - P1 presentation on Oct 15
  - Can be done in groups of 1-3 people. **Score multiplier: 1.1x for solo, 1.05x for two people and 1.0x for three people group.**
  - Project 1 is worth 20% credit.
    - You will be graded based on Understanding & Content (40%), Presentation & Communication (30%), Critical Analysis & Discussion (30%), by the teacher, TAs, and peers.

Grading - Machine Perception Homework

- Project 2: Machine Perception Homework (20%)
  - You will do an object detection homework.
  - Can be done in groups of 1-3 people.
  - You will submit to a public leaderboard. You will be evaluated based on model performance
  - Details to come. **Score cutoff, multiplier etc.**
  - Tentative schedule:
    - P2 & P3 release together on Oct 11
    - Due Dec 3
    - Top-3 team will be invited to present their methods on Dec 3. They get 4%/2%/1% extra credit.

Grading – Navigation Homework

- Project 3: Navigation Homework (30%)
  - You will do a social navigation homework.
  - Can be done in groups of 1-3 people.
  - You will submit to a public leaderboard. You will be evaluated based on model performance
  - Details to come. **Score cutoff, multiplier etc.**
  - Tentative schedule:
    - P2 & P3 release together on Oct 11
    - Due Dec 3
    - Top-3 team will be invited to present their methods on Dec 3. They get 4%/2%/1% extra credit.

Course Policy

- Late Policy

  Homework assignments and course project results are worth full credit on the due date. Unless granted an extension in advance, it is worth at most 75% credit for the next 48 hours, at most 50% credit after that. If you need an extension, please ask for it as soon as the need for it is known. Extensions that are requested promptly can be granted more liberally. You must turn in all assignments.

- Collaboration

  We encourage collaboration between students and studying materials in groups when the purpose of this is to facilitate learning, not to circumvent problems. It is allowed to seek help from other students in understanding the material needed to solve a particular problem. However, students must submit individual material and solutions, unless otherwise specified. Students should declare any collaboration on the first page of homework assignments (or equivalently on exercises). If the instructors believe the collaboration is improper, your grade may be affected

- Use of LLM/VLMs (ChatGPT, Gemini, etc.)
  - We encourage you to use ChatGPT to solve the homework assignment problems (because likely in the future everyone will use it for coding). However, you will need to declare using ChatGPT/LLMs in the homework README, as well as the prompts you use
  - **You are not allowed to use LLM for in-class test!!!**

Course Resources

- Textbooks
  1) Ian Goodfellow, Yoshua Bengio and Aaron Courville. Deep Learning. MIT Press, 2005.
  2) Sebastian Thrun, Wolfram Burgard, and Dieter Fox, Probabilistic Robotics, MIT Press, 2005.
  3) Introduction to Autonomous Mobile Robots - Illah Reza Nourbakhsh and Roland Siegwart
  4) Chinese textbook: 具身智能原理与实践

- Other Reading Materials
  - Nice talks
    - A Path Towards Embodied AI - Yann LeCun [link]
    - Robotics Professor Answers Robot Questions From Twitter - Henny Admoni [link]
  - Other Open Courses
    - Stanford AA 274a
    - CMU 16-311, 11-851
    - Washington CSE 478
    - Texas CS 391R

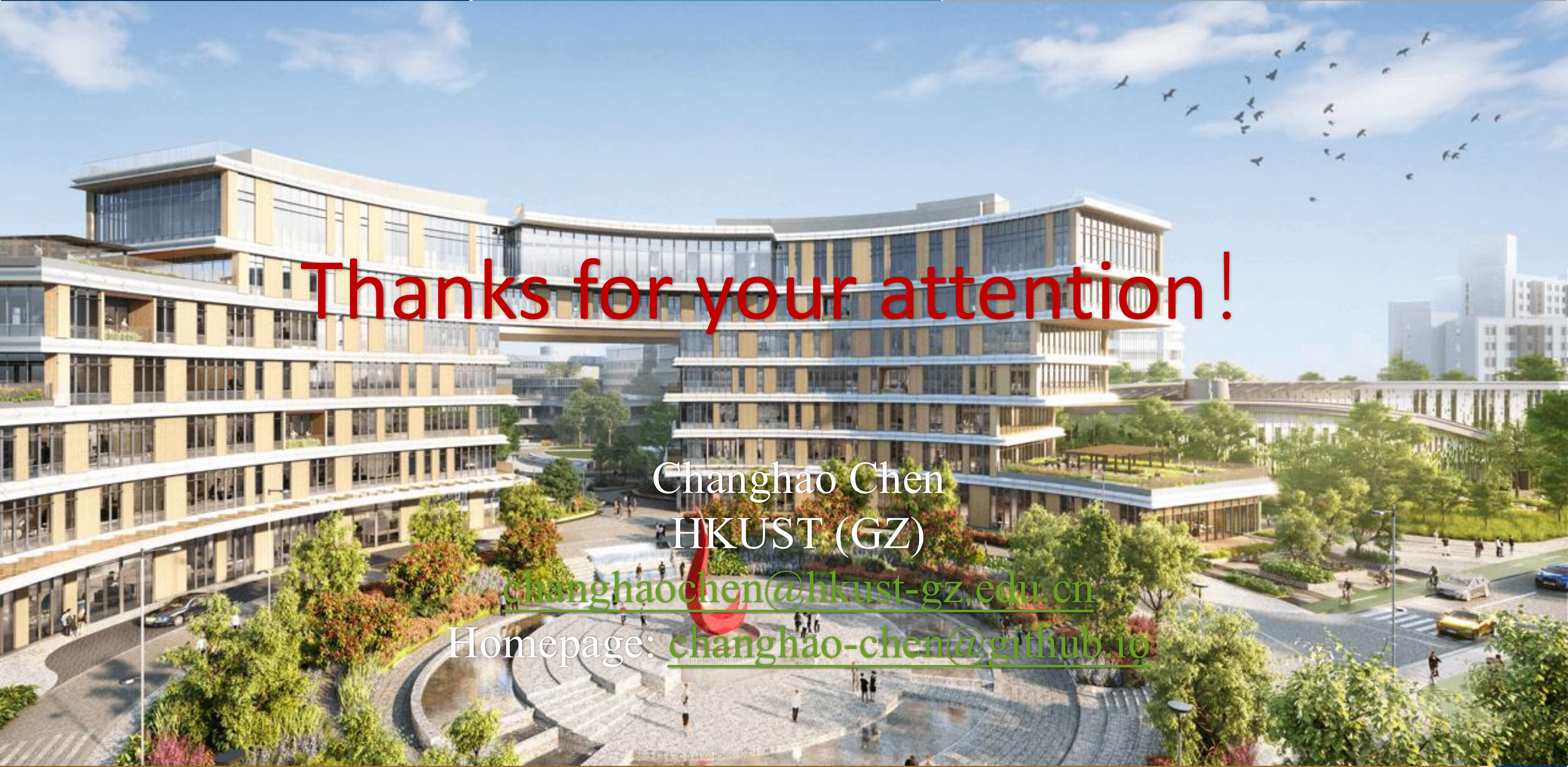
Important Resources for Reading

Journals:

- IEEE Transactions on Robotics
- The International Journal of Robotics Research
- IEEE Robotics and Automation Letters
- Journal of Field Robotics
- Science Robotics

Conferences:

- Robotics: Science and Systems
- IEEE International Conference on Robotics and Automation
- IEEE/RSJ International Conference on Intelligent Robots and Systems
- Computer Vision and Pattern Recognition
- International Conference on Computer Vision
- Neural Information Processing Systems



Thanks for your attention!

Changhao Chen
HKUST (GZ)
changhaochen@hkust-gz.edu.cn
Homepage: changhao-chen@github.io