16-867

Human Robot Interaction

Introduction

Instructor: Andrea Bajcsy

Carnegie Mellon University

intent ROBOTICS LAB

Welcome!

Professor

Andrea Bajcsy

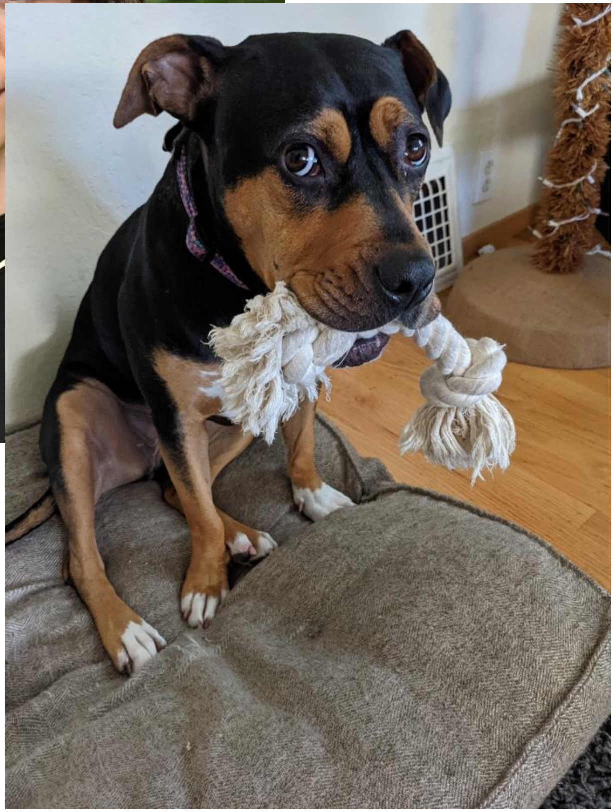
(BYE-chee)

What to call me:

- Andrea *(if you are a grad student)*
- Prof. Bajcsy or Prof. B *(if you are undergrad)*

Office Location: NSH 4629
Office Hours: Tuesdays, 12:20-1:20pm *(after class)*
Email: abajcsy@cmu.edu

Professor



Andrea Bajcsy

Fun fact: I have a dog named Cheerio…
she likes to sit like a human sometimes!

Teaching Assistant

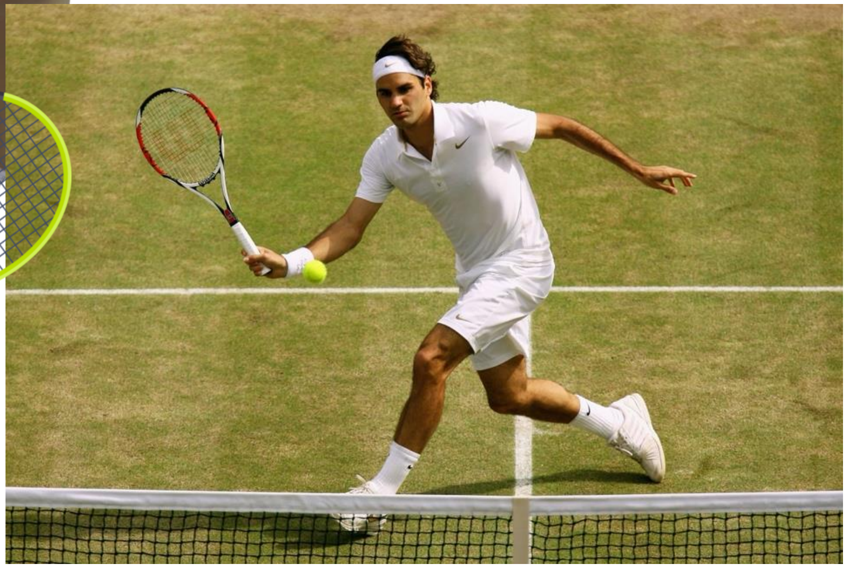
Yilin Wu, PhD Student

Research Interests:
- open-world, learning-based manipulation
- human-robot alignment

Office Location: NSH TBA
Office Hours: TBA
- **please take survey on Canvas so we can select OHs that suit folks best!**

Email: yilinwu@andrew.cmu.edu

Teaching Assistant



Yilin Wu, PhD Student

Fun fact: I love tennis! I'm a big fan of Roger Federer, but I only started playing tennis after he retired

What is next?

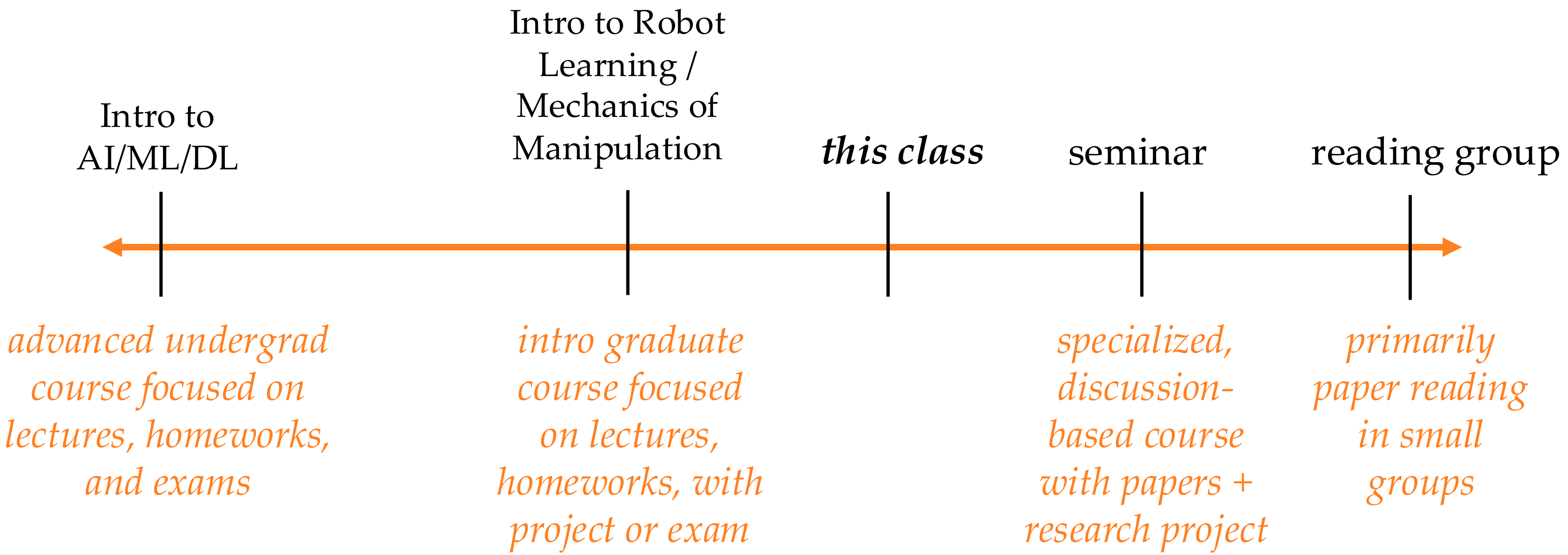
Course Content

Logistics

Intro Survey

(Intro to Single-Agent Decision Making)

What exactly is this class?



What makes human-robot interaction different from "typical" robotics?

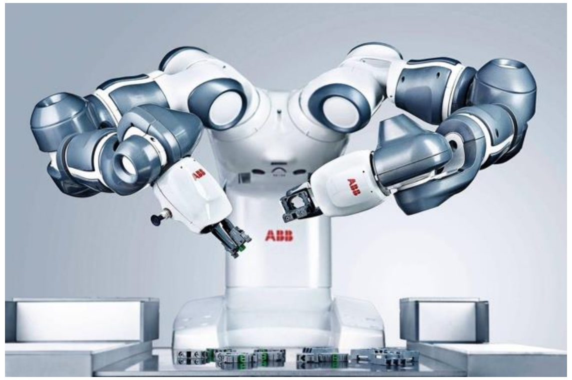
Boston Dynamics

ABB

Small group activity (5 min)

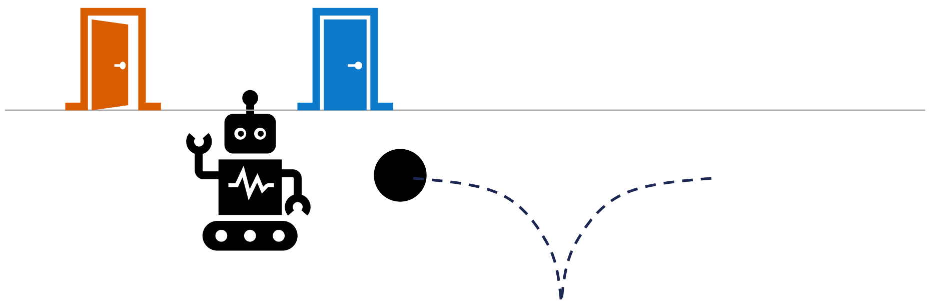
Turn to your neighbor(s), introduce yourself, and discuss:

What makes human-robot interaction different from "typical" robotics?



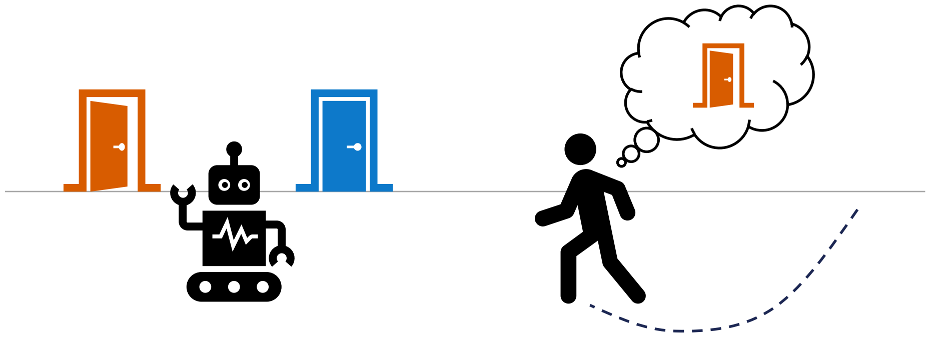
vs.

Interaction



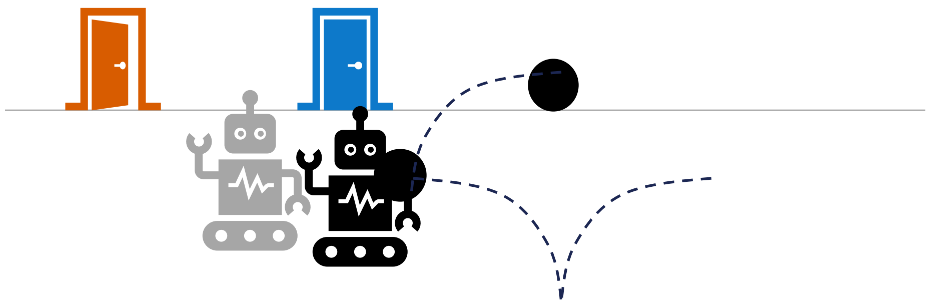
Environment driven by
laws of physics

Human Interaction



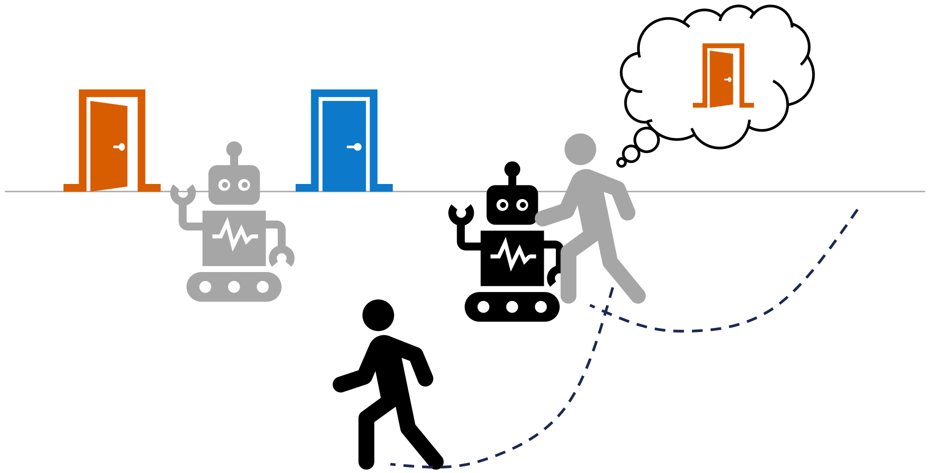
Human driven by physics and hidden
internal objectives

Interaction



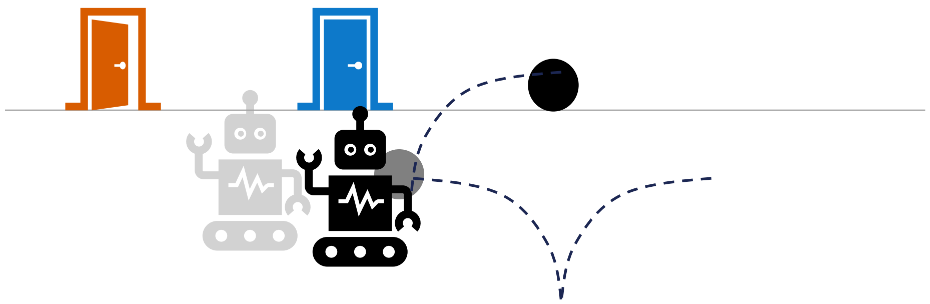
Environment can *be influenced* by robot's actions **directly**

Human Interaction



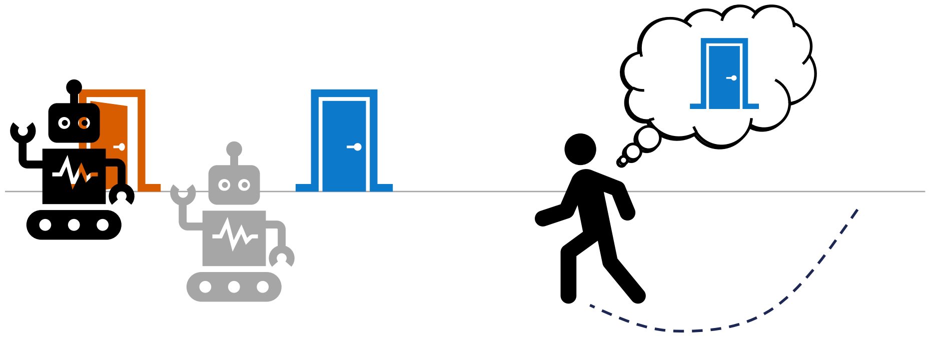
Human can *be influenced* by the robot's actions **directly….**

Interaction



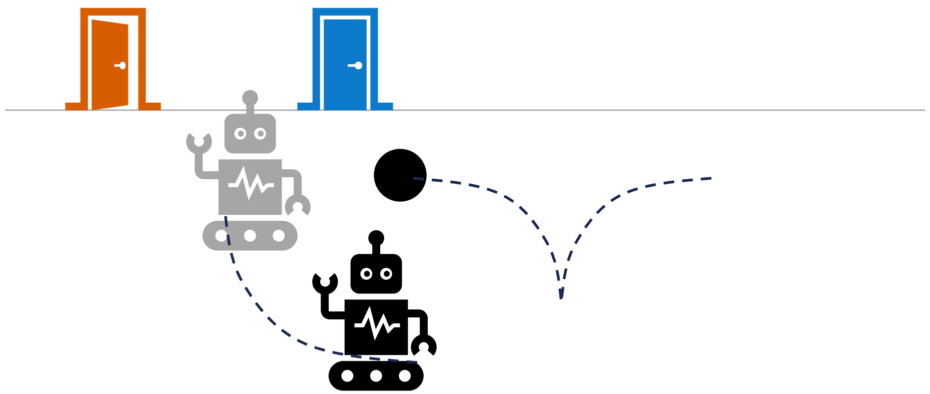
Environment can *be influenced* by robot's actions **directly**

Human Interaction



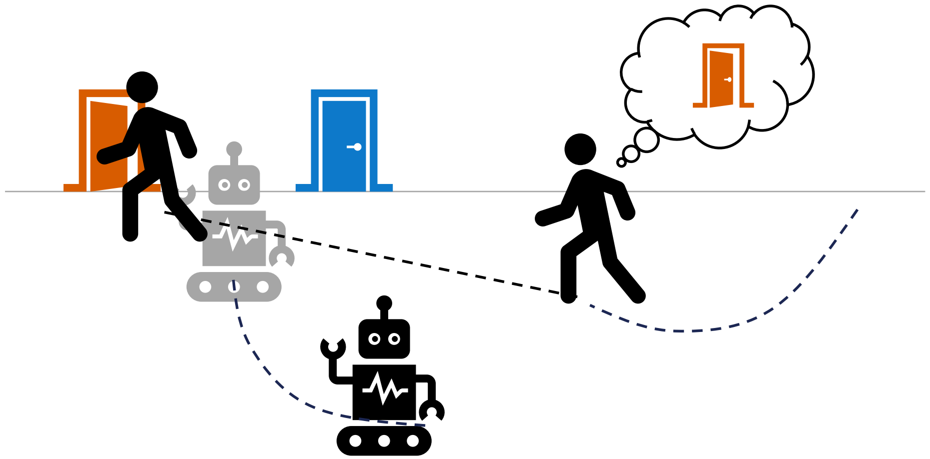
Human can *be influenced* by the robot's actions .... and **indirectly**

Interaction



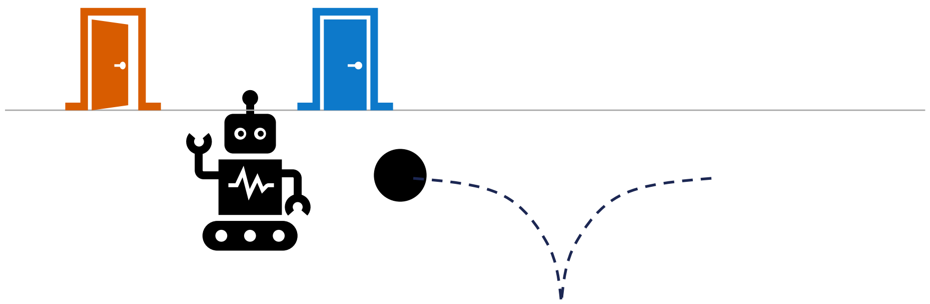
Environment can *influence* the robot's actions **indirectly**

Human Interaction



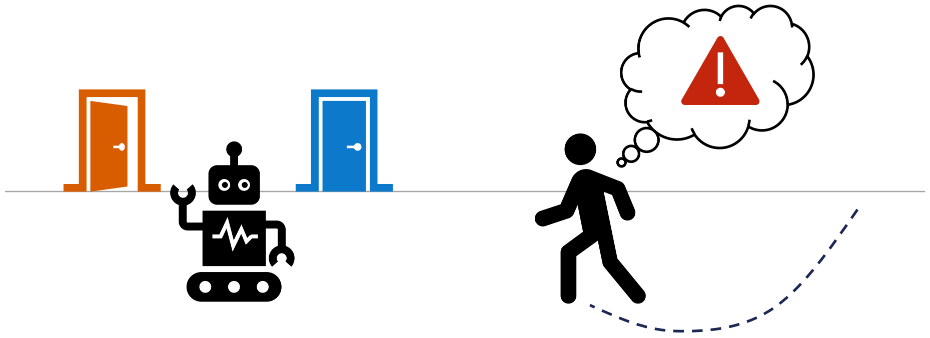
Human can *influence* the robot's behavior **directly or indirectly**

Interaction



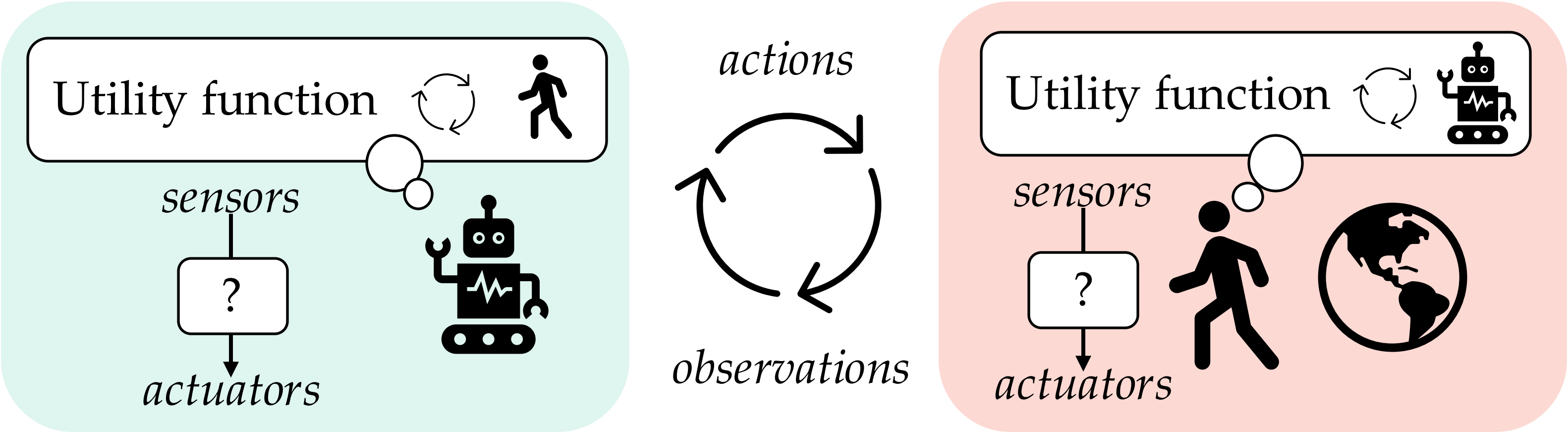
"Environment" is not a stakeholder

Human Interaction



Human is a stakeholder! (e.g., wants to derive value from robot)

How is algorithmic HRI different?



	See	Think	Act
Robotics	Detect objects, colors, estimate distances, poses, geometry, etc.	Plan / optimize for own utility function	Execute motor commands (e.g., based on physics models)
HRI	Detect objects, colors, distances, poses, geometry, etc. Estimate human's utility, intent, preferences, strategy/style	Plan / optimize for • Own utility function • Adapting / coordinating with human • Influencing human belief about the world or robot	May adjust execution dynamically based on human feedback (e.g., force feedback corrections)

But this seems really hard to encode into our algorithms… 🤔

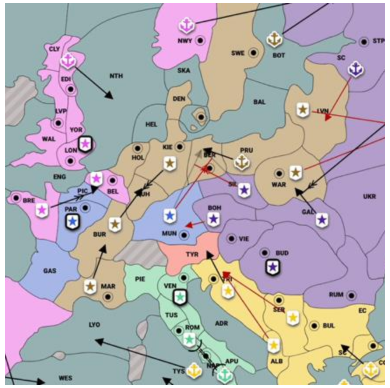
Where are people interacting with advanced autonomy the *most* right now?

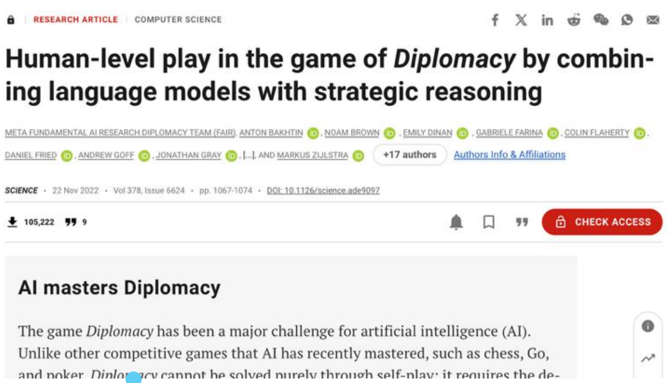
🤔

Exciting time for *interactive* Artificial Intelligence!

AlphaGo

CICERO




RESEARCH ARTICLE | COMPUTER SCIENCE

Human-level play in the game of *Diplomacy* by combining language models with strategic reasoning

META FUNDAMENTAL AI RESEARCH DIPLOMACY TEAM (FAIR), ANTON BAKHTIN, NOAM BROWN, EMILY DINAN, GABRIELE FARINA, COLIN FLAHERTY, DANIEL FRIED, ANDREW GOFF, JONATHAN GRAY, [...], AND MARKUS ZIJLSTRA +17 authors Authors Info & Affiliations

SCIENCE • 22 Nov 2022 • Vol 378, Issue 6624 • pp. 1067-1074 • DOI: 10.1126/science.ade9097

105,222 9 CHECK ACCESS

AI masters Diplomacy

The game *Diplomacy* has been a major challenge for artificial intelligence (AI). Unlike other competitive games that AI has recently mastered, such as chess, Go, and poker, *Diplomacy* cannot be solved purely through self-play; it requires the de-

Gemini

Supercharge your creativity and productivity

gemini.google.com

ChatGPT

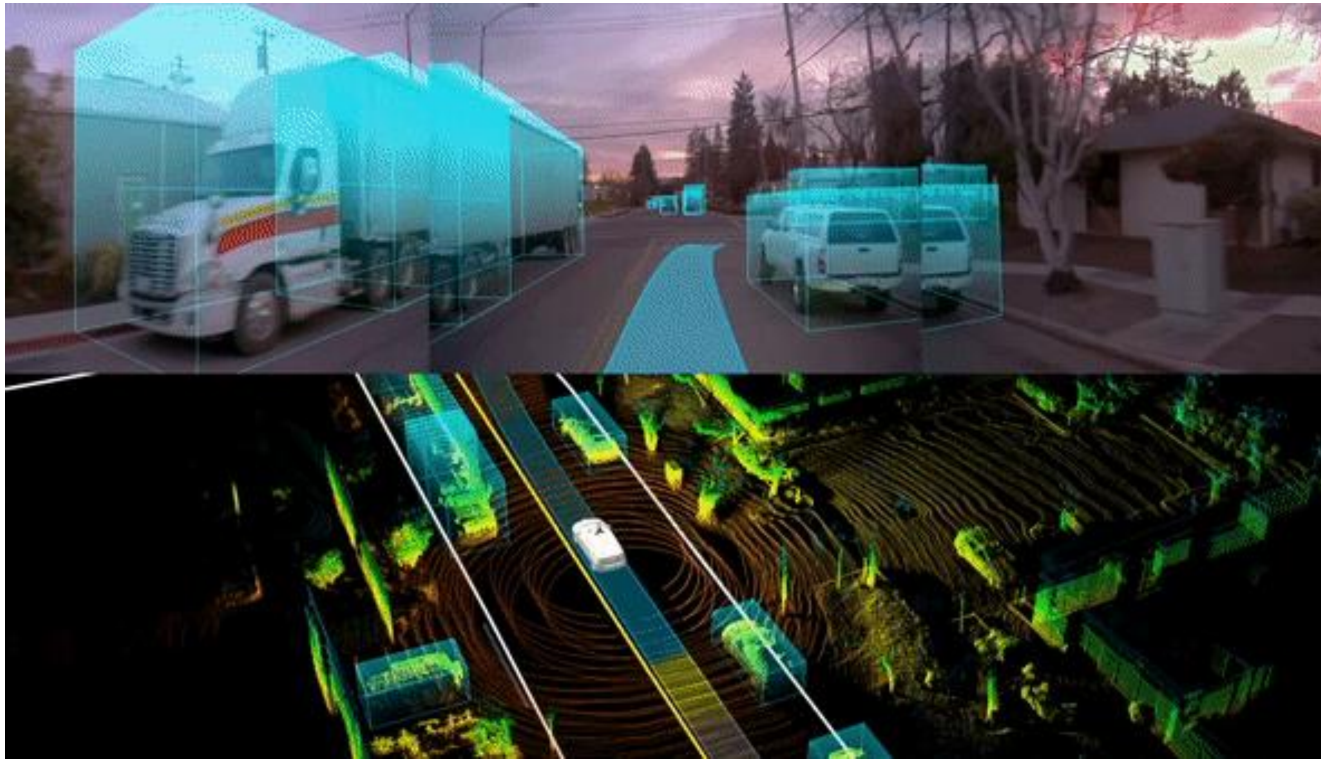
2016

2022

today

But where are the interactive **robots**?

How do robots interact with people today?



WAYMO

nuro

How do robots interact with people today?

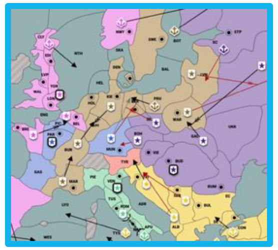
How do robots interact with people today?

Skydio

How do robots interact with people today?

hello robot™

Here is where we are in AI…



Gemini

ChatGPT

Fully robot-autonomous

Autonomy

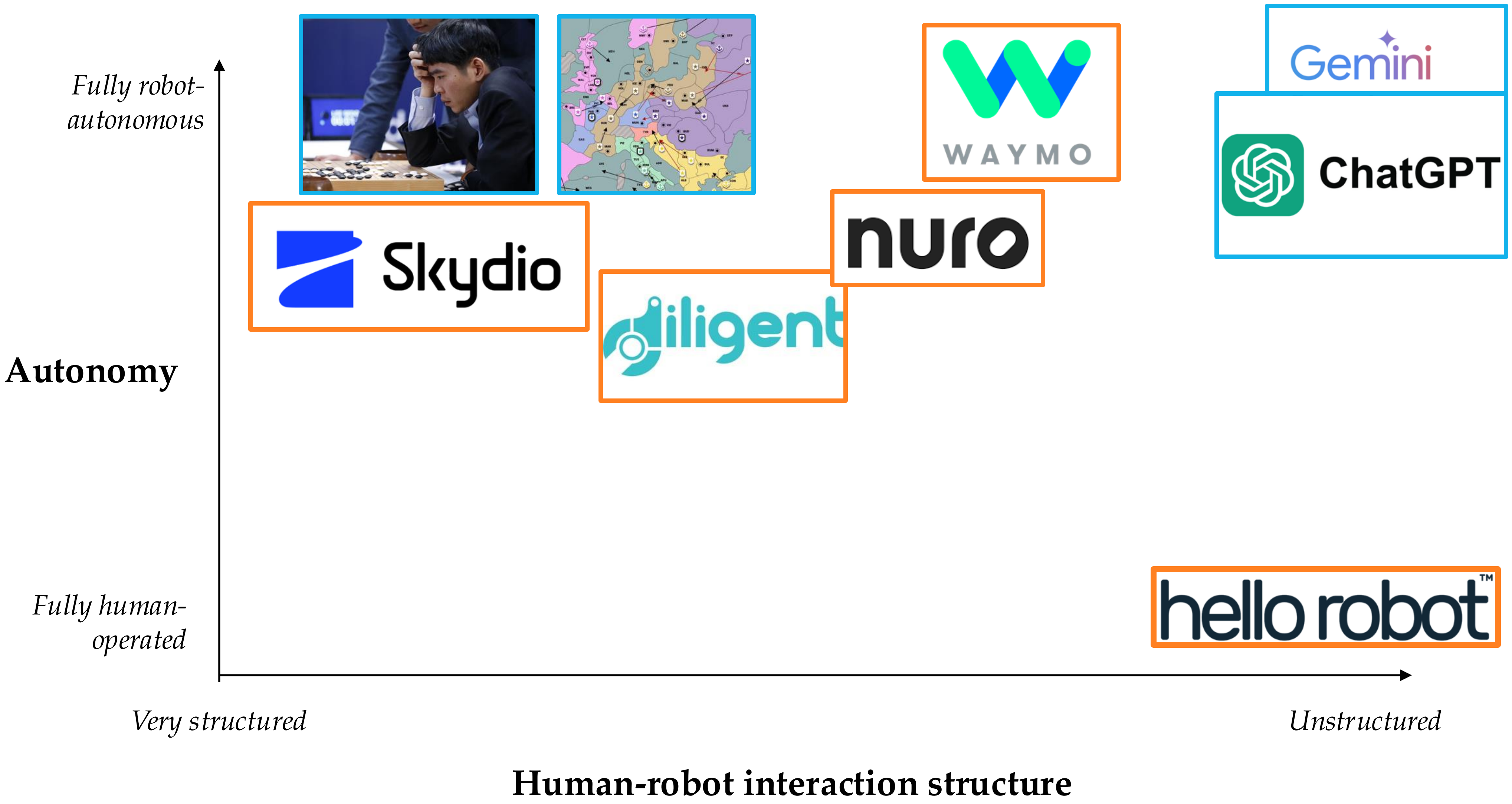
Fully human-operated

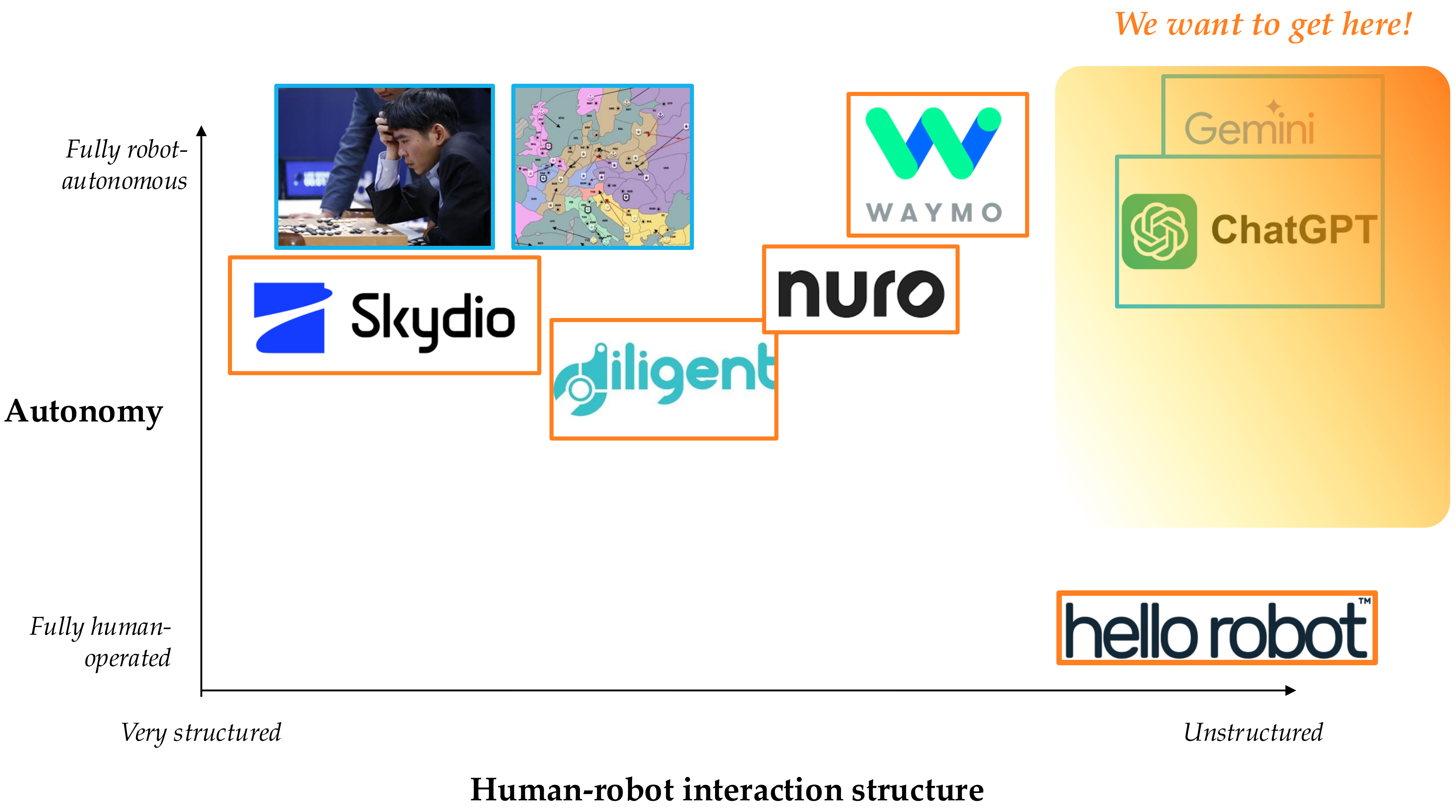
Very structured

Unstructured

Human-robot interaction structure

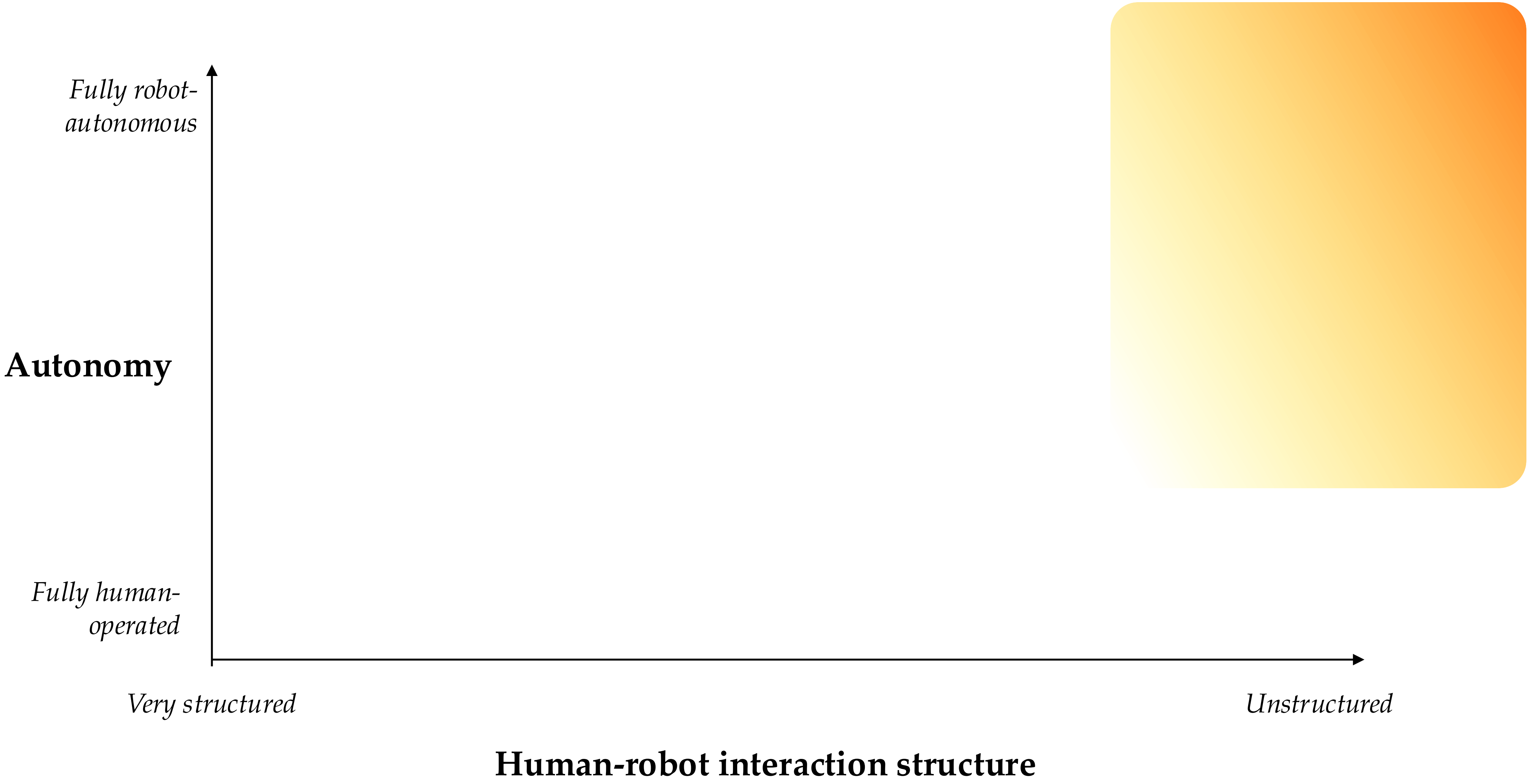
Here is where we are in robotics…





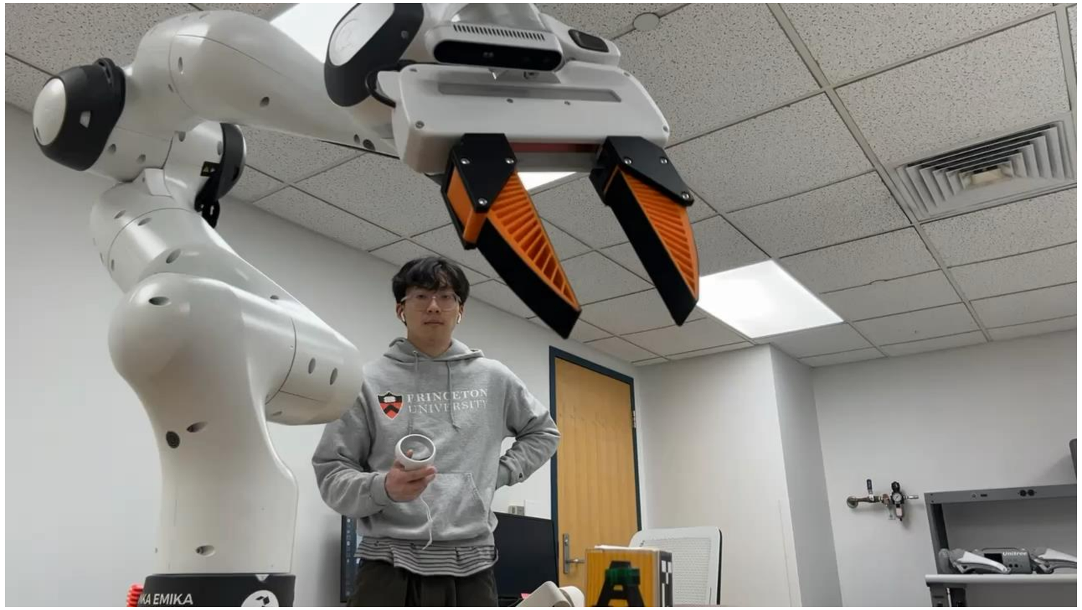
We want to get here!
Fully robot-autonomous
Autonomy
Fully human-operated
Skydio
diligent
nuro
WAYMO
Gemini
ChatGPT
hello robot
Very structured
Unstructured
Human-robot interaction structure

Think: Why are robots not in millions of homes?



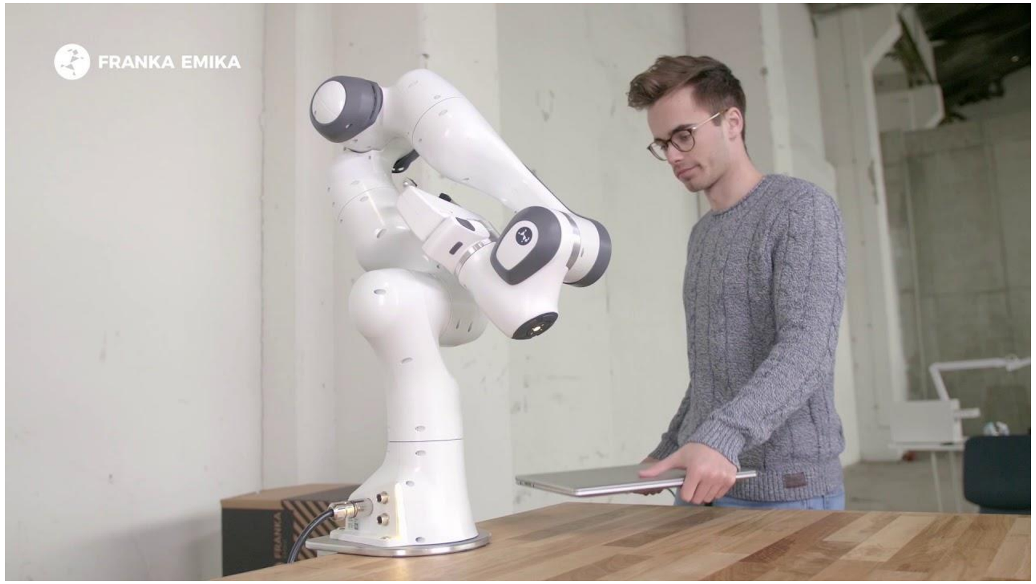
1. The way we program robots is rigid

Not flexible enough to be used by everyday users for everyday tasks; requires expert knowledge



Engineers Design Behaviors

Ship robot →



Users can't easily expand capabilities, or experience unexpected failures!

2. Hard to write down what "matters" to people

- ***Autonomy:*** *hard to design robot policies that behave according to what end-users want*
- ***Evaluation:*** *hard to write metrics that correlate with what end-users want*

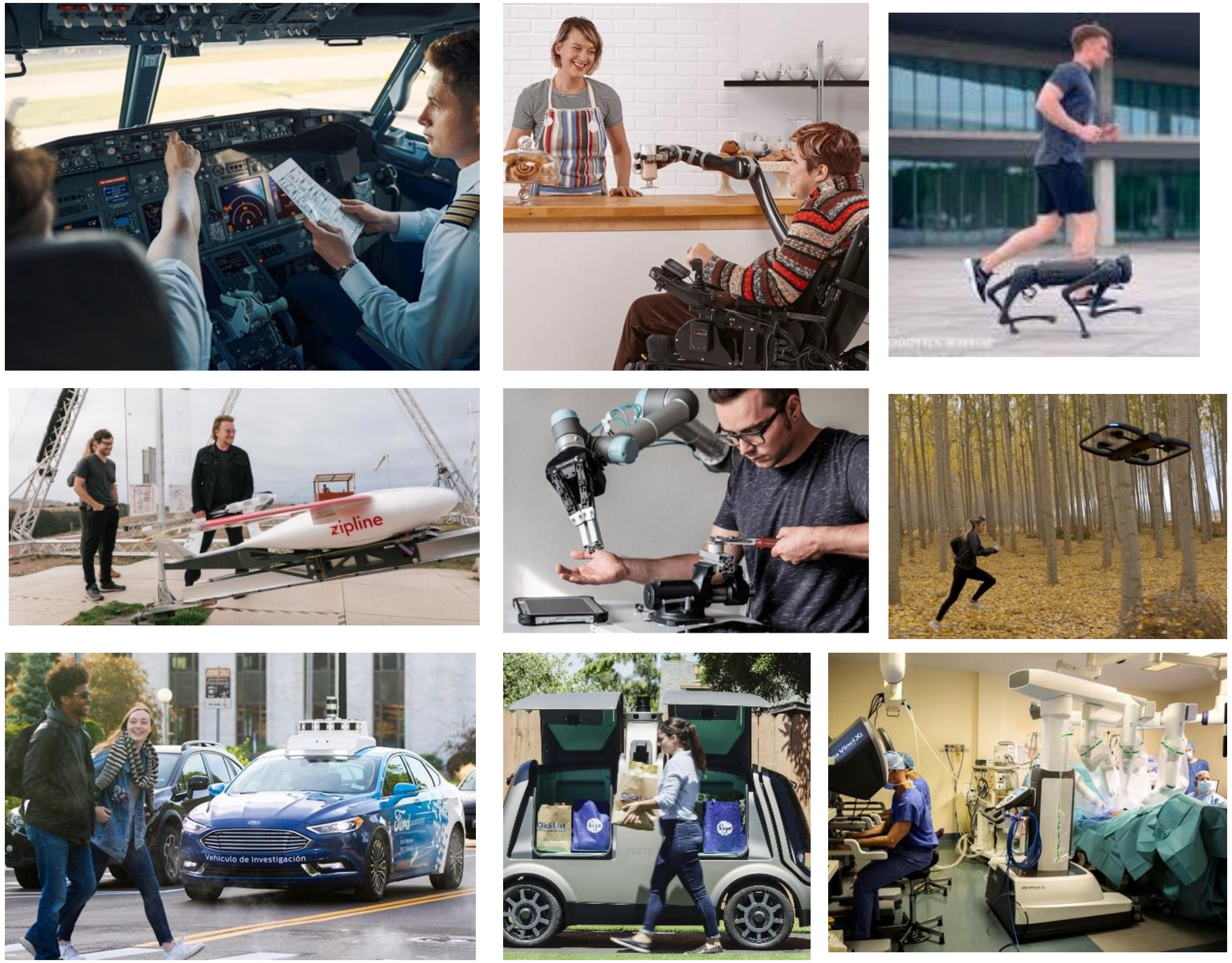
"Feel the Bite: Robot-Assisted Inside-Mouth Bite Transfer using Robust Mouth Perception and Physical Interaction-Aware Control." Jenamani, et al. (2024)

3. Hard to model human interaction

Human behavior is diverse: varying between individuals, environments, and over time

Why this course?

Take any robot application and …



1) **Model / quantify** human interaction with robots

2) **Solve** robot decision-making algorithms that are informed of/by people

3) **Identify** the frontiers of human-robot interaction

What you will learn in this course

Foundations

Single & multi-agent decision-making (MDPs, POMDPs, RL)
Intent inference & expression
Reward & policy learning
Experimental design

Prediction for Action

Trajectory forecasting
Collaboration, assistance, shared autonomy
Game theory

Learning, Alignment, and Safety

Alignment
Representation learning
HRI in Era of Foundation Models
Active learning
Safety & uncertainty quantification

Guest Lectures

Trajectory Forecasting

Ingrid Navarro
Senior PhD Student @ CMU

HRI in the Era of Foundation Models

Sidd Karamcheti
Toyota Research Institute |
Incoming Prof @ Georgia Tech

Active Learning

Erdem Biyik
Prof @ USC

General Resources

- No textbook!
- If I were to recommend textbooks for this class…

Artificial Intelligence: A Modern Approach by Russell and Norvig
Reinforcement Learning by Sutton and Barto
Probabilistic Robotics by Thrun, Burgard, Fox
Dynamic Noncooperative Game Theory by Başar and Olsder
Humans and Automation by Sheridan

Course Logistics

Format: lecture or related paper reading discussions

Typical 80-min class:

5-10 min logistics and recap
70 min lecture, invited talk, or paper discussion

Use *course website* for up-to-date schedule & paper links

https://abajcsy.github.io/human-robot-interaction/

Human Robot Interaction

Fall 2025. 16-867. Tuesday / Thursday 11:00am-12:20pm.

Announcements

Hello!

Nov 12 · 0 min read

See you next semester! 🤓

Course Overview

Robot interaction with people is inevitable: human engineers iteratively tune robot policies, autonomous cars nav crowded cities, construction workers teleoperate drones for building inspections, and assistive robots help end- daily living tasks.

In this graduate class, we will formalize such human-robot interaction (HRI) problems algorithmically. We will bui

Schedule (Tentative)

Foundations

Aug. 26:	**Course Overview**	Syllabus
Aug. 28:	**Single-Agent Sequential Decision-Making**	
Sept. 2:	**Value Iteration, Reinforcement Learning**	
Sept. 4:	**POMDPs, Probability, Bayesian Inference**	Goal Inference as Inverse Planning
Sept. 9:	**Intent Inference & Expression**	
Sept. 11:	**Intent Inference & Expression**	PAPER READING Expressing Thought, Functional Expressive Motion
Sept. 16:	**Reward and Policy Learning**	
Sept. 18:	**Experimental Design & Statistical Analysis**	HW #1 DUE A Primer for Conducting Experiments in Human–Robot Interaction

Prediction for Action

Sept. 23:	GUEST LECTURE **Trajectory Forecasting (Ingrid Navarro, CMU)**	
Sept. 25:	**Trajectory Forecasting**	PAPER READING Confidence-Aware Prediction, ManiCast
Sept. 30:	**Collaboration, Assistance, & Coordination**	

Please read the *Syllabus* on the *course website*

https://abajcsy.github.io/human-robot-interaction/

Syllabus

TABLE OF CONTENTS

Overview

Robot interaction with people is inevitable: human engineers iteratively tune robot policies, autonomous cars navigate through crowded cities, construction workers teleoperate drones for building inspections, and assistive robots help end-users with daily living tasks. In this class we will formalize such human-robot interaction problems algorithmically. We will build the mathematical foundations for modeling human-robot interaction across robots and tasks, enable robots to understand human intent and predict human behavior, and study how robot learning

Also: Helpful answers to ***FAQs***!

https://abajcsy.github.io/human-robot-interaction/

FAQ

TABLE OF CONTENTS

I don't have access to robots! What can I do?

If you want to use **physical hardware**, consider using the AI Maker Space in Tepper for your projects! To get more info, you can contact the AI Maker Space manager: Greg Armstrong at ai-makerspace@cs.cmu.edu.

If using physical hardware is infeasible for your project, consider using **simulation** or **benchmark datasets** for controlled testing and algorithm design. Some relevant resources include:

Autonomous Driving

- CARLA - driving simulator with urban layouts, buildings, vehicles, and sensors
- CARLO - a lightweight 2D version of CARLA

Use *Canvas* for downloading / uploading assignments

16867-A

Fall 2025

- **Home**
- Announcements
- Syllabus
- Assignments
- Quizzes
- Grades
- Discussions
- Files
- People
- Zoom
- NameCoach
- Syllabus Registry
- Pages
- Outcomes
- Collaborations
- Rubrics
- Modules

Recent Announcements

Human Robot Interaction

Assign To | Edit

Welcome to **16-867: Human-Robot Interaction!**

Robot interaction with people is inevitable: human engineers iteratively tune robot policies, autonomous cars navigate through crowded cities, construction workers teleoperate drones for building inspections, and assistive robots help end-users with daily living tasks.

In this graduate class, we will formalize such human-robot interaction (HRI) problems algorithmically. We will build the mathematical foundations for modeling human-robot interaction across robots and tasks, enable robots to understand human intent and predict human behavior, and study how robot learning changes in the presence of human feedback. The approaches covered will draw upon a variety of disciplines and tools such as sequential decision-making, cognitive science, Bayesian inference, and modern machine learning. Throughout the class, there will also be several guest lectures from experts in the field. Students will practice essential research skills including reviewing papers, writing project proposals, and technical communication.

Grading

*See **class syllabus on course website** for detailed info*

Percentage	Activity
10%	Attendance & Participation
20%	HW (2x)
10%	Paper Summaries
5%	Project Proposal
20%	Midterm Project (Report + Presentation)
35%	Final Project (Report + Presentation)

Attendance & Participation (10%)

Expected to attend class in person—this is how we will all get the most out of the class! <u>Please show up on time</u>, especially for reading days

The way we grade this:

- <u>First 5 minutes of class:</u> we will give a **short, easy "quiz"** related to the last lecture's content. This is graded as 1/0.
  - e.g., *"Describe what is a sequential decision-making problem."*
- **Permitted 2 unexcused absences**, no questions asked, before being docked.

I understand that occasionally you may have challenges attending (e.g., illness, religious observance,..); **please let me know.**

Homework (20%)

HW #1: Intent Inference & Expression

Released:	~Sept 4
Due:	Sept 18

HW #2: Human-Robot Alignment

Released:	~Nov. 4
Due:	Nov. 13

These are coding-based homeworks in **Python** and **PyTorch**. They are *not* meant to be tedious; they are meant to **empower** you! ☺

If you are not confident (or are rusty) with Python and Pytorch, please come see us for educational resources!

Paper Summaries (10%)

Prediction for Action

Sept. 23:	GUEST LECTURE	Trajectory Forecasting (Ingrid Navarro, CMU)	
Sept. 25:		Trajectory Forecasting	PAPER READING Confidence-Aware Prediction, ManiCast
Sept. 30:		Collaboration, Assistance, & Coordination	
Oct. 2:		Shared Autonomy	PAPER READING Shared Autonomy via Hindsight Optimization, LILA
Oct. 7:		HRI as a Game	
Oct. 9:		HRI as a Game	PAPER READING Planning for AVs that Effect Humans, Long-term Robot Influence on Humans

Paper discussion days:

8 paper reading days
2 papers per reading day

Before class:

Answer three questions about the paper:

1) What assumptions were made about (a) robot, (b) human, (c) their interaction
2) What extensions would you propose? Scope & justify
3) What do you like about this work?

Must submit on Canvas before class.

In class:

Split you into small groups, discuss set of questions, I assign a representative from each group to present on the group's takeaways, and the whole class can engage on the answer

Class Project

Two options:

Research project:

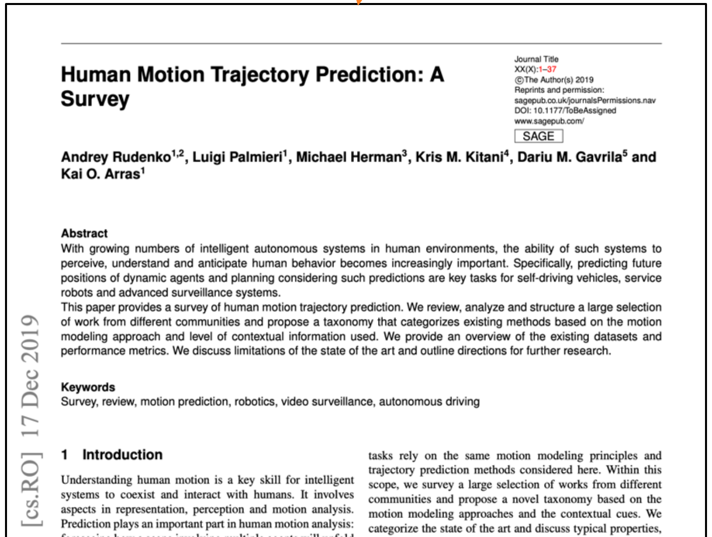
Identify a research direction broadly relevant to this class
Propose and take first steps towards an original idea

Literature survey:

Select a topic area and rigorous way in which you will find papers
Characterize this topic area in an <u>insightful way</u> (e.g., open questions, common assumptions, tractable vs. theoretical gaps)

You must work in a group of min 2 to max 5 people. Explicit permission from Andrea must be granted if you want to work on an independent project

Example of good literature survey



Human Motion Trajectory Prediction: A Survey

Journal Title
XX(X):1–37
©The Author(s) 2019
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Andrey Rudenko[1,2], Luigi Palmieri[1], Michael Herman[3], Kris M. Kitani[4], Dariu M. Gavrila[5] and Kai O. Arras[1]

Abstract
With growing numbers of intelligent autonomous systems in human environments, the ability of such systems to perceive, understand and anticipate human behavior becomes increasingly important. Specifically, predicting future positions of dynamic agents and planning considering such predictions are key tasks for self-driving vehicles, service robots and advanced surveillance systems.
This paper provides a survey of human motion trajectory prediction. We review, analyze and structure a large selection of work from different communities and propose a taxonomy that categorizes existing methods based on the motion modeling approach and level of contextual information used. We provide an overview of the existing datasets and performance metrics. We discuss limitations of the state of the art and outline directions for further research.

Keywords
Survey, review, motion prediction, robotics, video surveillance, autonomous driving

[cs.RO] 17 Dec 2019

1 Introduction

Understanding human motion is a key skill for intelligent systems to coexist and interact with humans. It involves aspects in representation, perception and motion analysis. Prediction plays an important part in human motion analysis:

tasks rely on the same motion modeling principles and trajectory prediction methods considered here. Within this scope, we survey a large selection of works from different communities and propose a novel taxonomy based on the motion modeling approaches and the contextual cues. We categorize the state of the art and discuss typical properties,

Class Project

When picking a project, make sure to answer the question:

How does the project connect to the broader topics & context of the class?

Examples of projects within scope (non-exhaustive list!)

- Applying one of the techniques from class to your problem domain
  - *(e.g., using game-theory for pedestrian prediction, using RLHF to align an assistive AI agent, developing a new shared controller)*
- Rigorously comparing two methods that seek to solve the same problem
  - *(e.g., RLHF vs. DPO for aligning a diffusion policy, running a pilot human study to compare 2 shared controller designs)*
- Posing (and solving) a new algorithmic HRI problem for your problem domain
- Challenging an assumption underlying one of the methods in the class

Class Project

When picking a project, make sure to answer the question:

How does the project connect to the broader topics & context of the class?

When in doubt: come talk to us about your interests and we can help!

⚠ Examples of projects not within scope

- No clear connection to human interaction
  - *e.g., a new self-supervised learning method for image classification; a robot motion planner for doing assembly tasks*
  - *To make them in scope: e.g., using human feedback to improve an image classification algorithm; robot motion planner which takes as input natural language descriptions of the assembly task*
- No clear connection to robotics / AI
  - *e.g., designing a new user interface for a mobile phone*
  - *To make them in scope: e.g., designing a new interface for humans to program robot behavior*

Class Project

Project Proposal (5%) – <u>due</u>: Sept. 30

1. **Report:** max 1 page project pitch

Mid-term Project (20%) – <u>due</u>: Oct 23

1. **Report:** max 4 page writeup of progress
2. **In-class Oral Presentation:** short conference-style project pitch (~3-5 mins)

Final Project (35%)

1. **Report:** max 6 page writeup of findings
   <u>due</u>: Dec 11
1. **In-class Oral Presentation:** short conference-style lightning talk pitch (~8-10 mins)
   <u>due</u>: Dec 1

Class Project

Project Proposal (5%) – **due: Sept. 30**

1. **Report:** max 1 page project pitch

Mid-term Project (20%) – **due: Oct 23**

1. **Report:** max 4 page writeup of progress
2. **In-class Oral Presentation:** short conference-style project pitch (~3-5 mins)

Final Project (35%)

1. **Report:** max 6 page writeup of findings
   due: Dec 11
1. **In-class Oral Presentation:** short conference-style lightning talk pitch (~8-10 mins)
   due: Dec 1

Project Proposal

Published | Assign To | Edit

This is a brief (maximum 1 page, excluding references) project pitch. You can think of this as an extended abstract: you want to motivate the topic you have chosen and answer some key brainstorming questions about your directions.

Please use the attached Latex template and answer the questions below (which are also in the template):

- Motivation: describe the context of your project. What is the setting / environment, tasks, or human-robot interaction you are considering, etc? Who do you think will be most interested in your project? What will your project enable in the future?
- Open Challenge(s): what is the core challenge (or challenges) you want to tackle? What makes your problem hard? What has been holding us back from solving this; i.e., why don't we have an answer to this yet?
- Proposed Approach: brainstorm some approaches you may take to tackle the challenges. why are these approaches promising or feasible? how will you measure success?
- Risks: what are some risks or roadblocks you anticipate?

Latex Template (zip file): project-proposal-latex.zip

Project Proposal: Your Project Title Here

Michael Shell
Email: mshell@ece.gatech.edu

Homer Simpson
Email: homer@thesimpsons.com

Marge Simpson
Email: marge@thesimpsons.com

I. MOTIVATION

(describe the context of your project. What is the setting / environment, tasks, or human-robot interaction you are considering, etc? Who do you think will be most interested in your project? What will your project enable in the future?)

II. OPEN CHALLENGE(S)

(what is the core challenge (or challenges) you want to tackle? What makes your problem hard? What has been holding us back from solving this; i.e., why don't we have an answer to this yet?)

III. PROPOSED APPROACH

(brainstorm some approaches you may take to tackle the challenges. why are these approaches promising or feasible? how will you measure success?)

IV. RISKS

(what are some risks or roadblocks you anticipate?)

REFERENCES

Class Project

Project Proposal (5%) – **due: Sept. 30**

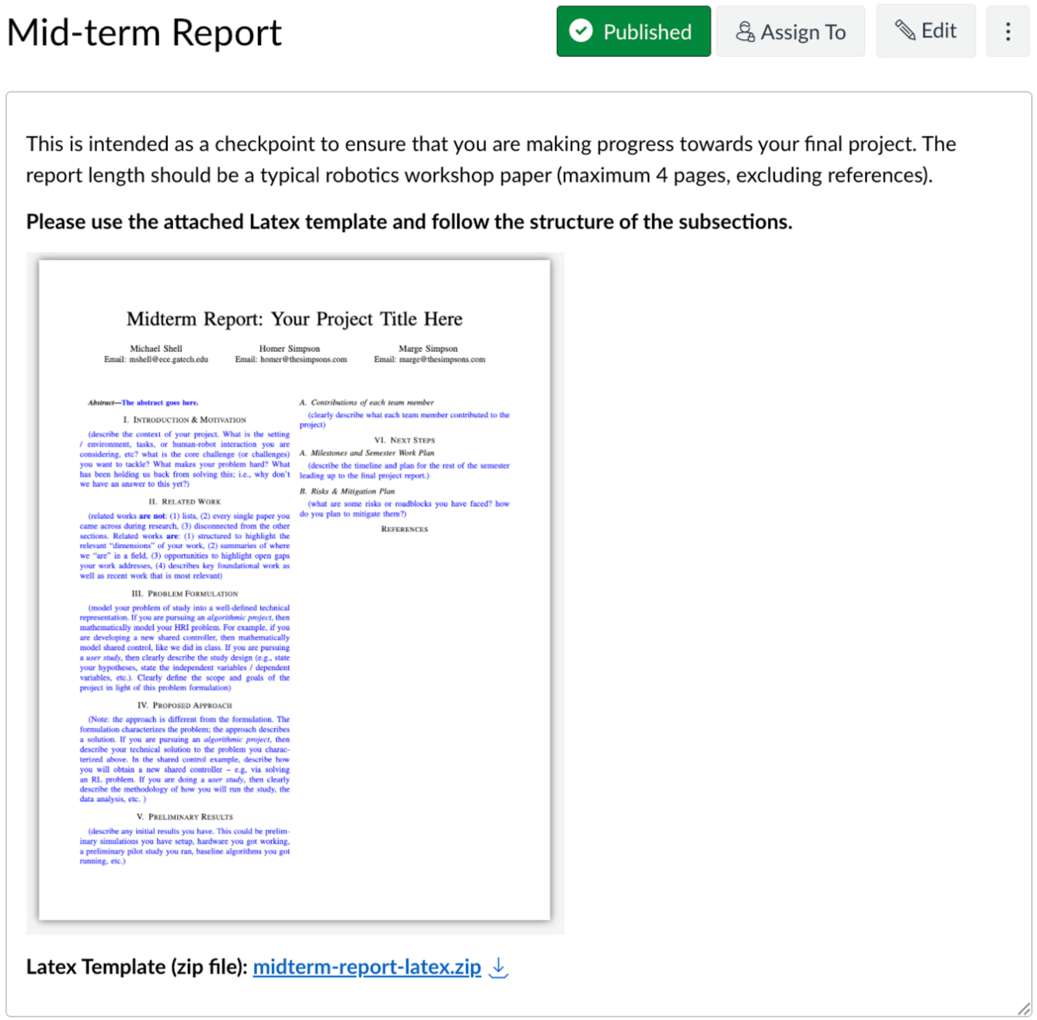
1. **Report:** max 1 page project pitch

Mid-term Project (20%) – **due: Oct 23**

1. **Report:** max 4 page writeup of progress →
2. **In-class Oral Presentation:** short conference-style project pitch (~3-5 mins)

Final Project (35%)

1. **Report:** max 6 page writeup of findings
   due: Dec 11
1. **In-class Oral Presentation:** short conference-style lightning talk pitch (~8-10 mins)
   due: Dec 1



Class Project

Project Proposal (5%) – <u>due</u>: Sept. 30

1. **Report:** max 1 page project pitch

Mid-term Project (20%) – <u>due</u>: Oct 23

1. **Report:** max 4 page writeup of progress
2. **In-class Oral Presentation:** short conference-style project pitch (~3-5 mins)

Final Project (35%)

1. **Report:** max 6 page writeup of findings
   <u>due</u>: Dec 11
1. **In-class Oral Presentation:** short conference-style lightning talk pitch (~8-10 mins)
   <u>due</u>: Dec 1

Final Project Report

Published | Assign To | Edit

The final report should present your final findings in a research or survey paper format. The length should be maximum 6 pages, double-column. The grade will be determined based on the content quality and not on the absolute length (please see the grading rubric below).

Please use the attached Latex template and follow the structure of the subsections.

Final Report: Your Project Title Here

Michael Shell
Email: mshell@ece.gatech.edu

Homer Simpson
Email: homer@thesimpsons.com

Marge Simpson
Email: marge@thesimpsons.com

Abstract—**The abstract goes here.**

I. INTRODUCTION & MOTIVATION

(describe the context of your project. What is the setting / environment, tasks, or human-robot interaction you are considering, etc? what is the core challenge (or challenges) you want to tackle? What makes your problem hard? What has been holding us back from solving this; i.e., why don't we have an answer to this yet?)

II. RELATED WORK

(related works **are not**: (1) lists, (2) every single paper you came across during research, (3) disconnected from the other sections. Related works **are**: (1) structured to highlight the relevant "dimensions" of your work, (2) summaries of where we "are" in a field, (3) opportunities to highlight open gaps your work addresses, (4) describes key foundational work as well as recent work that is most relevant)

III. PROBLEM FORMULATION

(model your problem of study into a well-defined technical representation. If you are pursuing an *algorithmic project*, then mathematically model your HRI problem. For example, if you are developing a new shared controller, then mathematically model shared control, like we did in class. If you are pursuing a *user study*, then clearly describe the study design (e.g., state your hypotheses, state the independent variables / dependent variables, etc.). Clearly define the scope and goals of the project in light of this problem formulation)

IV. PROPOSED APPROACH

(Note: the approach is different from the formulation. The formulation characterizes the problem; the approach describes a solution. If you are pursuing an *algorithmic project*, then describe your technical solution to the problem you characterized above. In the shared control example, describe how you will obtain a new shared controller – e.g. via solving an RL problem. If you are doing a *user study*, then clearly describe the methodology of how you will run the study, the data analysis, etc.)

V. RESULTS

(describe the results you have. This includes, but is not limited to, **qualitative** results of your simulations/hardware/pilot study experiments like robot trajectories, human predictions, human expressions, etc.; **quantitative** results like bar/line plots comparing baselines to the propose methods in terms of relevant metrics, results of statistical tests like ANOVAs, Likert Scale survey results, etc.)

VI. CONCLUSION

(summarize what the problem was and what was learned in this project)

A. Future Work

(describe the limitations of this work and what more there is to be done)

B. Contributions of Each Team Member

(clearly describe what each team member contributed to the project)

REFERENCES

Latex Template (zip file): final-report-latex.zip

Survey (5 min)

https://forms.gle/SPqv62u1EyC9SFHt9

16-867

Human Robot Interaction

Introduction

Instructor: Andrea Bajcsy

Carnegie Mellon University

intent ROBOTICS LAB