# CMSC 671 Principles of Artificial Intelligence:
(Guest Lecture)

# Planning and Logical Agent Ethics

October 10, 2023

Cassandra Kent, PhD

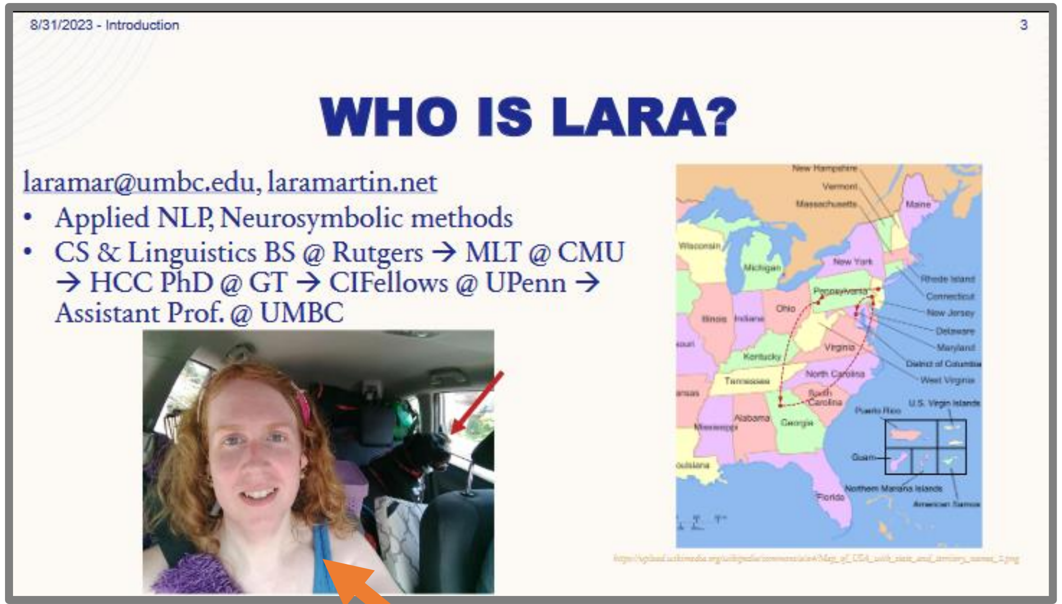
Some material adapted from Dr. Cynthia Matuszek

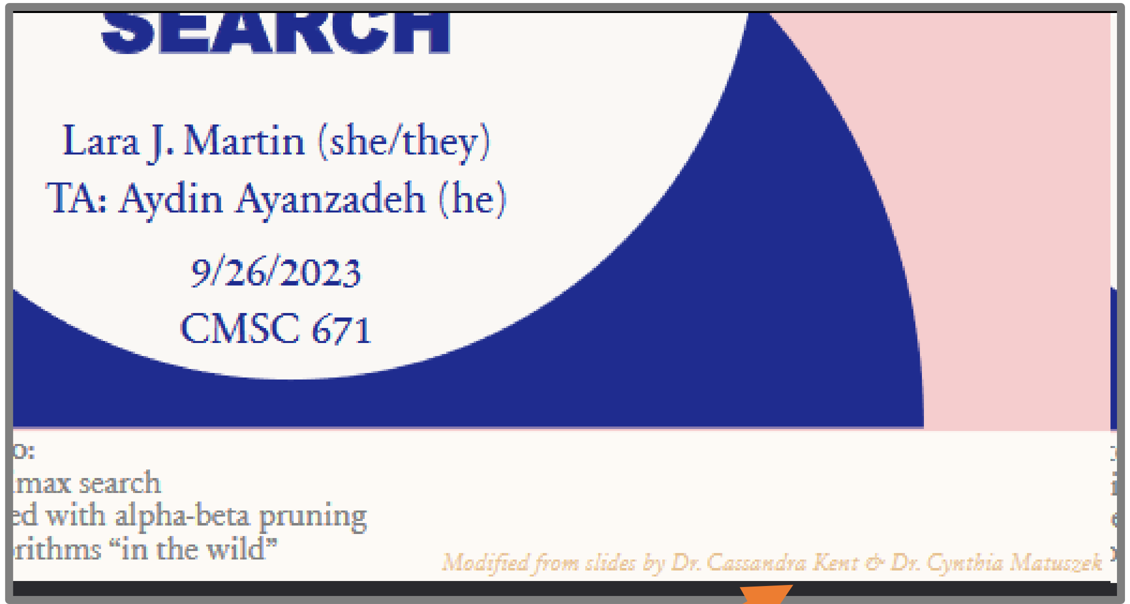
By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# But first… Who is your guest lecturer?

I’m not Dr. Martin, but for the purposes of this class I’m probably the next closest person.




**(This is me)** **(Also me)**

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# But first… Who is your guest lecturer?

- My name is Cassandra Kent, my pronouns are she/her
- I have a PhD in robotics from Georgia Tech, and recently completed a postdoc at UPenn
- As a teacher, I’m a certified CIRTL associate through GT’s Tech to Teaching program

# But first… Who is your guest lecturer?

## Research

- I do robotics research in human-robot interaction (HRI), robot learning from demonstration, collaborative robotics, and lifelong (machine) learning - more details at my scholar page and website

- I'm both an **AI researcher** and an **AI practitioner**

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# But first… Who is your guest lecturer?

Other considerations:

- I just completed my postdoc and am currently unaffiliated with an institution - taking time to recover from long COVID complications

Why am I telling you this?

- You can reach me after these lectures through Dr. Martin
- Please be patient with word-finding issues
- If I can't come up with a specific term, I will make a note and get back to you after class
- We'll take a short break about halfway through the class

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# CMSC 671 Principles of Artificial Intelligence:
## (Guest Lecture)
# Planning and Logical Agent Ethics

October 10, 2023

Cassandra Kent, PhD

Some material adapted from Dr. Cynthia Matuszek

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Introduction to Planning

## What is planning (in AI)?

- Combination of search and logic to solve complex problems
- Often used in modern AI research and applications (compare/contrast with reinforcement learning - next module)

Automated story generation

AIIDE

Artificial Intelligence for Interactive Digital Entertainment Conference

Human-robot collaboration

By the end of class today, you will be able to:

1. **Extend ideas from logic to define a planning problem for an agent with a task**
2. **Extend ideas from search to solve planning problems**
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Planning Problem

- What is the planning problem?
- Find a **sequence of actions** that achieves a **goal** when executed from an **initial state**.
- That is, given
  - A set of operators (possible actions)
  - An initial state description
  - A goal (description or conjunction of predicates)
- Compute a sequence of operations: a **plan**.

By the end of class today, you will be able to:

1. **Extend ideas from logic to define a planning problem for an agent with a task**
2. **Extend ideas from search to solve planning problems**
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Let's define a planning problem…

Planning problem for a robot doing making a sandwich

*literals from propositional or first-order logic*

**State representation**: b : bread ^ v:vegetables ^ c:cheese ^ k: knife; chopped(v: vegetables) ^ …

**Initial state:** unchopped(v) ^ k ^ b ^ c ^ clean(k) ^ …

**Goal**: E(c, b1, b2, v) On(c, b1) ^ On(v, c) ^ On(b2, v) ^ chopped(v) ^ chopped(b) ^ …

*define as **action schema** – includes **preconditions** and **effects***

**Actions**:

chop(f: food, k: knife)

preconditions: unchopped(f) ^ clean(k) ^ …

**effects**: add-list: chopped(f) ^ …

delete-list: clean(k)

By the end of class today, you will be able to:

1. **Extend ideas from logic to define a planning problem for an agent with a task**
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Standardization of Planning Problems

## Assumptions of Classical Planning

Environment assumptions:

- **Fully observable** – agent has complete knowledge of the state
- **Deterministic** – no uncertainty in results of actions
- **Static** – only the agent can cause change in the world
- **Finite** and **Discrete**

Other typical assumptions:

- **Atomic time** – each action is indivisible
- **No concurrent actions**
- **Closed world assumption**
  - state descriptions list true literals
  - anything not listed is known false

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Standardization of Planning Problems

## Assumptions of **Classical Planning**

Environment assumptions:

Other typical assumptions:

Allows for standardization of planning languages:
**Planning Domain Definition Language** (**PDDL**) – includes STRIPS, ADL, etc.

- **Finite** and **Discrete**

- anything not listed is known false

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Solving Planning Problems

Now that we've defined a planning problem, **how do we find a solution**?

Prompts:

1. How do the search approaches you know map onto planning?
2. Why are we adding this new structure? What benefits do we get?
3. Through the lens of search, what challenges do you anticipate when solving planning problems?

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. **Extend ideas from search to solve planning problems**
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Solving Planning Problems - Search-based Approach

**We have everything we need for search** – initial state, goal, actions, ~~transition model~~, ~~uniform step cost~~

encoded in action schema — typical assumption

- **Progression planner**: search forward
  - From initial state to goal state
  - Directly analogous to search
- **Regression planner**: search backward
  - From goal state to initial state
  - Works iff actions have enough information to go both ways
  - Ideally leads to reduced branching: planner is only considering things that are relevant to the goal

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. **Extend ideas from search to solve planning problems**
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Solving Planning Problems - Search-based Approach

- Problem with treating planning as search – potentially have very large branching factors
- As with search, we need a good **heuristic for planning problems**
- **Let's talk about constructing admissible heuristics for general planning problems**

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. **Extend ideas from search to solve planning problems**
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Planning Heuristics

Approach 1: derive heuristics by solving **relaxed planning problems**

- Remove all action preconditions - every action applies in any state
- Remove all negative effects (i.e. delete lists) - actions only move towards the goal
- Use state abstraction - replace instances in a literal with more general classes

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. **Extend ideas from search to solve planning problems**
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Planning Heuristics

Approach 2: Assume **subgoal independence**

- Find plan cost for individual literals in a goal conjunction
- Add up each individual cost for the final heuristic
- *Note: not always admissible! E.g., one action may achieve multiple subgoals at the same time*

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. **Extend ideas from search to solve planning problems**
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Planning Heuristics

Approach 3: Take advantage of useful pre-computed data structures

- Compute a **planning graph** to construct heuristics - see Chapter 11.4 in Russell & Norvig for more details
- Retrieve exact solution costs of common sub-problems stored in a database

Approach 4: Heuristic composition - combine any of the previous approaches

**Need to manage tradeoffs between heuristic computation time vs. how effective they are at speeding up the search process.**

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. **Extend ideas from search to solve planning problems**
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems

# Why use planning?

Take about 1 minute and write down some **advantages** of using the planning approaches we discussed today to solve real-world problems.

Take about 1 minute and write down some **disadvantages** of using the planning approaches we discussed today to solve real-world problems.

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. **Assess the advantages and disadvantages of planning**
4. Reflect on ethical considerations when applying logical agents to real problems

# Why use planning?

## Advantages

- a lot information describing the world
- gives structure for goal definition
- we know how it will behave

## Disadvantages

- can miss things in planning specification
- many many assumptions
- intractability
- hard or time-consuming to specify that much info
- not flexible if something changes

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. **Assess the advantages and disadvantages of planning**
4. Reflect on ethical considerations when applying logical agents to real problems

# Planning Wrap-Up

- Planning combines search and logic to create specific formalisms useful for solving complex sequential problems
- Once a problem is specified, there are many approaches to find a solution: Search-based, Situation Calculus*, Partial-Order Planning*, Satisfiability Planning*, … *See Russell & Norvig Chapter 11 for implementation details!
- The best approach typically depends on characteristics of the planning domain – there's no general-purpose best method!
- Standardization in planning problem definition, such as PDDL, lets us develop many problem-agnostic approaches to try

By the end of class today, you will be able to:

1. **Extend ideas from logic to define a planning problem for an agent with a task**
2. **Extend ideas from search to solve planning problems**
3. **Assess the advantages and disadvantages of planning**
4. Reflect on ethical considerations when applying logical agents to real problems

# AI Ethics: Logical Agents

Let's return to our **question-based framework** for identifying possible ethical issues with AI approaches

First, we'll discuss the following prompt:

*Comparing to search, do logical agents (propositional, first-order, and planning-based) introduce any new* ***societal implications*** *or* ***potential issues****?*

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. **Reflect on ethical considerations when applying logical agents to real problems**

# AI Ethics: Logical Agents

*Comparing to search, do logical agents (propositional, first-order, and planning-based) introduce any new **societal implications** or **potential issues**?*

- protecting the users if they're vulnerable
- issues with the knowledge base, making sure that biases aren't coming from the knowledge base itself
- easy to overlook details

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. **Reflect on ethical considerations when applying logical agents to real problems**

# AI Ethics: Logical Agents - Applying our Previous Questions

*Explainability and Interpretability*

- How clear is our agent's decision making? Is it transparent or is it a black box?
  - Example: Logical agents can express their reasoning - see HW2
- Can we make changes to the algorithm to make its decisions more explainable?
  - Example: Logical agents can return their reasoning with their actions - again see HW2
- Can we develop tools that make the algorithm's decisions easier to interpret?
  - Example: Rules to translate from FOL to natural language

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. **Reflect on ethical considerations when applying logical agents to real problems**

# AI Ethics: Logical Agents - Applying our Previous Questions

*Job Displacement*

- Will this algorithm displace human workers?
  - Example: Planning agents could replace: healthcare workers, service industries
- Will this algorithm/agent create new jobs? Who will benefit?
  - Example: We need knowledge base engineers to build up initial axioms - this heavily favors people with CS or mathematics educations

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. **Reflect on ethical considerations when applying logical agents to real problems**

# AI Ethics: Logical Agents - Applying our Previous Questions

*Inequality*

- Who has access to this AI agent?
  - Could this create new inequality between groups that have access and do not have access?
- Is this system reinforcing existing structures that create inequality?
  - If yes, is there regulation for this technology that can prevent this?

*We'll come back to this in a moment…*

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. **Reflect on ethical considerations when applying logical agents to real problems**

# AI Ethics: Logical Agents - New Questions

*Data Concerns*

**This will come back in a big way in the ML module!**

- Can this approach introduce data privacy concerns?
  - Example: consider shared knowledge bases across personal agents, medical agents, etc.
  - If yes, are there policies or approaches to mitigate these privacy concerns?
    - Example: for building KBs - data anonymization, opt-in user data sharing, etc.
- What biases might be present in the data itself?
  - Are there policies or approaches to address these biases?
    - Example: user-centered design to construct the knowledge base
- Do we have access to the data required to develop this agent/model? How do we know?
  - Example: extract common-sense axioms from Wikipedia text under a Creative Commons license

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. **Reflect on ethical considerations when applying logical agents to real problems**

# AI Ethics: Logical Agents - Applying our Previous Questions

*Inequality*

- Who has access to this AI agent?
  - Could this create new inequality between groups that have access and do not have access?
- Is this system reinforcing existing structures that create inequality?
  - Example: extract axioms from public text - can encode negative stereotypes, learn slurs, etc.
  - If yes, is there regulation for this technology that can prevent this?
    - Example: require moderation for knowledge bases built from data mining

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. **Reflect on ethical considerations when applying logical agents to real problems**

# What to do before next class (Thursday, October 12)

- Read Sections 16.1-16.3, 17.1-17.2
  - There's nothing to submit today or Thursday
  - MDPs are foundational to RL, this is very relevant to modern AI approaches!

- Homework is not due this week!
- (but think about making some progress on it anyway)

- I'll get these slides to Dr. Martin to post on the course website

By the end of class today, you will be able to:

1. Extend ideas from logic to define a planning problem for an agent with a task
2. Extend ideas from search to solve planning problems
3. Assess the advantages and disadvantages of planning
4. Reflect on ethical considerations when applying logical agents to real problems