

CSCE 580: Introduction to AI

Week 6 - Lectures 11 and 12: AI Trust; Symbolic - Representation and Logic

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

23RD AND 25TH SEP 2025

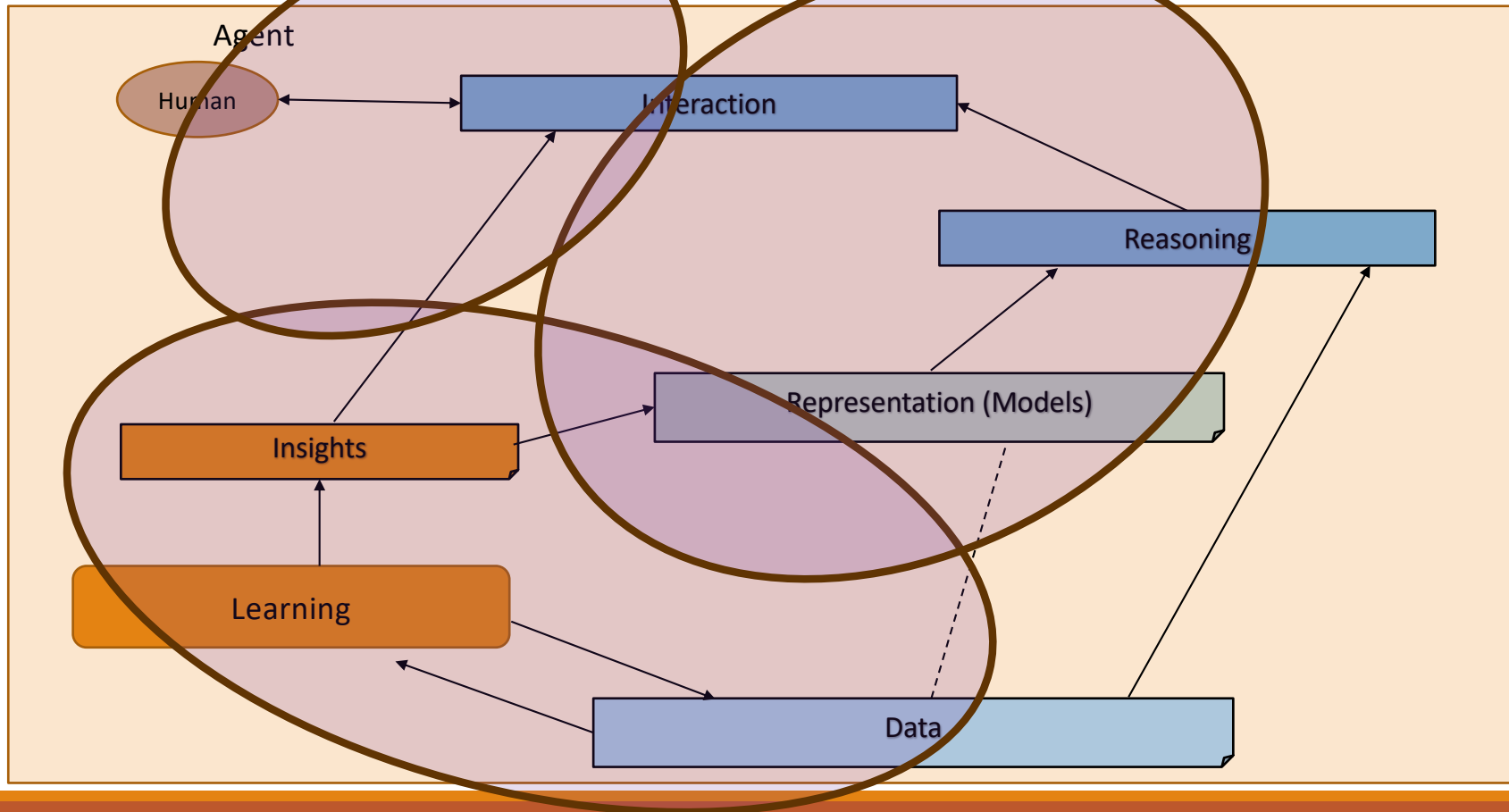
Carolinian Creed: “I will practice personal and academic integrity.”

Credits: Copyrights of all material reused acknowledged

Organization of Week 6 - Lectures 11, 12

- Introduction Section
 - Recap from Week 5 (Lectures 9 and 10)
 - AI news
- Main Section
 - L11: ML Trust Issues – Explainability, Rating
 - L12: Symbolic - Representation and Logic
- Concluding Section
 - About next week – W7: Lectures 13, 14
 - Ask me anything

Relationship Between Main AI Topics (Covered in Course)



Recap of Week 5

- We talked about
 - Language models
 - LLMs
 - Using LLMs
 - AI/ LLM Trust
 - Exercise 1, Project A

- Week 1: Introduction, Aim: Chatbot / Intelligence Agent
- Weeks 2-3: Data: Formats, Representation and the Trust Problem
- Week 3: Machine Learning – Supervised (Classification)
- Week 4: Machine Learning - Unsupervised (Clustering) –
- Topic 5: Learning neural network, deep learning, Adversarial attacks
- Week 6: Large Language Models – Representation and Usage issues
- Weeks 7-8: Search, Heuristics - Decision Making
- Week 9: Constraints, Optimization – Decision Making
- Topic 10: Markov Decision Processes, Hidden Markov models -
Decision making
- Topic 11-12: Planning, Reinforcement Learning – Sequential decision making
- Week 13: Trustworthy Decision Making: Explanation, AI testing
- Week 14: AI for Real World: Tools, Emerging Standards and Laws; Safe AI/ Chatbots

AI News

#1 NEWS – Discovering Plants Around Us (SC), Virtually!

- Link: <https://herbarium.org/> - Initiative started by Dr. Andrew Charles Moore over a century back of dried plant specimens

The screenshot shows the SERNEC (Southeast Regional Network of Expertise and Collections) website. The header includes the SERNEC logo and a 'Sign In' button. The navigation bar lists: Home, Specimen Search +, Images +, Inventories +, Dynamic Tools +, Data Use, and Sitemap. Below the navigation bar, the breadcrumb trail reads: Home >> Search Criteria >> Specimen Records. The main content area has three tabs: Species List (selected), Occurrence Records, and Maps. Under the Species List tab, the dataset is identified as 'USCH-' and the search criteria as 'has images; excluding cultivated/captive occurrences'. A pagination bar shows '1 2 3 4 5 6 7 8 9 10 11 >> Last' and 'Page 1, records 1-100 of 104847'. Three specimen records are displayed, each with a small thumbnail image of the plant specimen:

- Symplocos tinctoria (L.) L'Hér.**
USCH:ACM ACM0367 Moore, A. s.n. 1926-04-05
United States, South Carolina, Orangeburg, Branchville
[Full Record Details](#)
- Iris cristata Aiton**
USCH:ACM ACM0027 Moore, A. 1925-04-12
United States, South Carolina, Richland, Ridgewood
[Full Record Details](#)
- Alnus incana subsp. rugosa (Du Roi) R.T. Clausen**
USCH:ACM ACM0124 Moore, A.
United States, South Carolina, Richland, Campus of University of South Carolina
[Full Record Details](#)

Potential usage

1. Learn about plants
2. Learn how they change with location and time
3. How fall colors change over time

Example

- <https://sernecportal.org/portal/collections/list.php?db=147&includeothercatnum=1&hasimages=1&usethes=1&taxontype=2&association-type=none&taxontype-association=2&usethes-associations=1&comingFrom=newsearch>

Introduction Section

Lecture 11: Overcoming ML Trust Issues – Explainability, Rating

Lecture 11: Concluding Comments

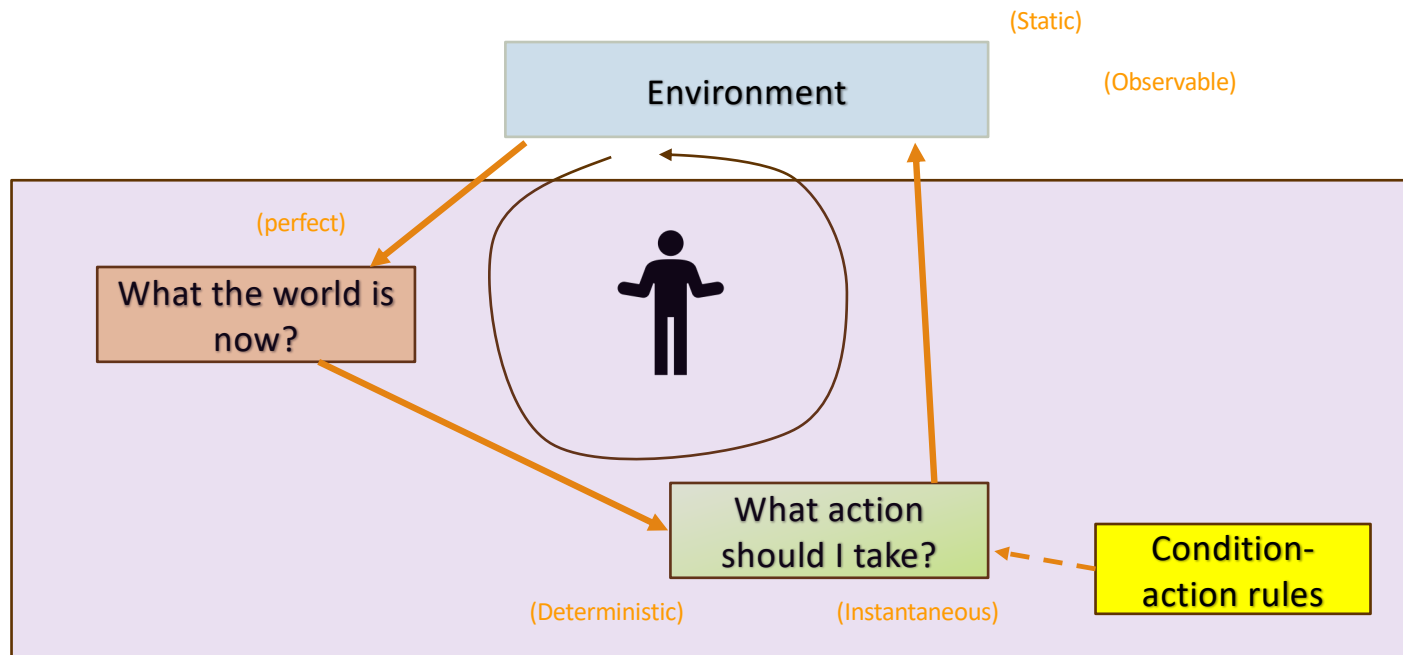
We discussed

- AI trust and risk mitigation
- Explainability methods
- Rating methods

Lecture 12: Representation and Logic

Main Section

Intelligent Agent – Simple Knowledge Based

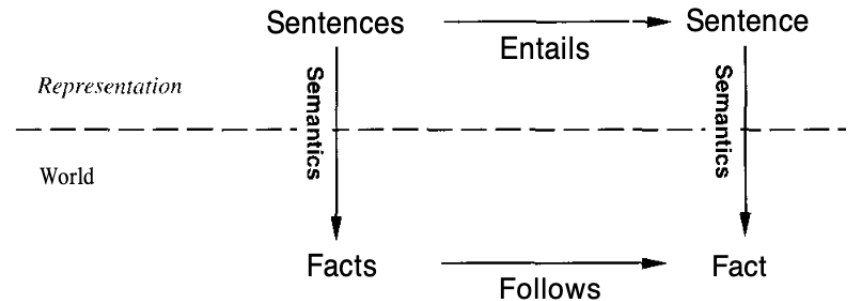


Logic – Basic Idea

- Starting with **true assumptions**, a knowledge-based system (automaton) to draw **true conclusions**
- Logic consists of three components
 - Syntax -- allowable sentences
 - Semantics -- determining truth of sentence given a model (assignment of values to sentences)
 - Inference Procedure -- rules to draw conclusion given a set of sentences

Formal Logic – 1/3

- An automaton for manipulating symbols and drawing conclusions
- Consists of a knowledge base with:
 - a set of true statements (sentences). Sentences have
 - Syntax
 - Semantics – compositional property
 - Proof theory: a set of rules for deducing the entailments / interpretations of the sentences
- Properties of sentences
 - **Valid:** A sentence is **valid** or necessarily true if and only if it is true under all possible interpretations in all possible worlds. Also called a **tautology**
 - **Satisfiable:** A sentence is satisfiable if and only if there is some interpretations in some possible worlds where it is true.



Credits:

- Russell & Norvig, AI - A Modern Approach
- Deepak Khemani - A First Course in AI

Propositional Logic

- Sentences: assertions about the world
 - Atomic sentence: propositional symbol
 - ClassToday – whether there is a class today: Yes, No
 - A -- any fact of interest: Yes, No
 - True -- always true
 - False -- always false

Propositional Logic - Syntax

Sentence — *AtomicSentence* | *ComplexSentence*

AtomicSentence — *True* | *False*

| *P* | *Q* | *R* | ...

ComplexSentence — (*Sentence*)

| *Sentence* *Connective* *Sentence*

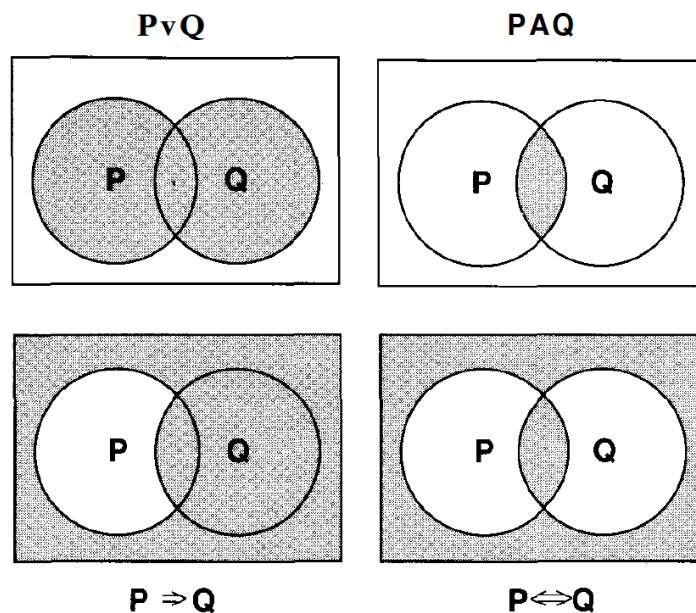
| \neg *Sentence*

BNF (Backus-Naur Form) grammar
of sentences in propositional logic

Connective — *A* | *V* | \Leftrightarrow | \Rightarrow

Source: Russell & Norvig, AI: A Modern Approach

Propositional Logic - Semantics



Model of sentence: Any world in which a sentence is true (under a particular interpretation)

α	β	\neg	$\alpha \vee \beta$	$\neg \beta \vee \neg$	$\alpha \vee \neg$
False	False	False	False	True	False
False	False	True	False	True	True
False	True	False	True	False	False
False	True	True	True	True	True
True	False	False	True	True	True
True	False	True	True	True	True
True	True	False	True	False	True
True	True	True	True	True	True

Truth table to prove soundness of inference

Source: Russell & Norvig, AI: A Modern Approach

Inference Procedure

- Given a knowledge base (KB), generate new sentences α that are entailed by KB
 - $KB \models \alpha$
- Given KB and α , report whether or not α is entailed by KB
 - $KB \models \alpha$
-

Propositional Logic

Inference Procedures

- ◇ **Modus Ponens or Implication-Elimination:** (From an implication and the premise of the implication, you can infer the conclusion.)

$$\frac{a \Rightarrow \beta, \quad a}{\beta}$$

- ◇ **And-Elimination:** (From a conjunction, you can infer any of the conjuncts.)

$$\frac{\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n}{\alpha_i}$$

- ◇ **And-Introduction:** (From a list of sentences, you can infer their conjunction.)

$$\frac{\alpha_1, \alpha_2, \dots, \alpha_n}{\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n}$$

- ◇ **Or-Introduction:** (From a sentence, you can infer its disjunction with anything else at all.)

$$\frac{\alpha_i}{\alpha_1 \vee \alpha_2 \vee \dots \vee \alpha_n}$$

- ◇ **Double-Negation Elimination:** (From a doubly negated sentence, you can infer a positive sentence.)

$$\frac{\neg\neg\alpha}{\alpha}$$

- ◇ **Unit Resolution:** (From a disjunction, if one of the disjuncts is false, then you can infer the other one is true.)

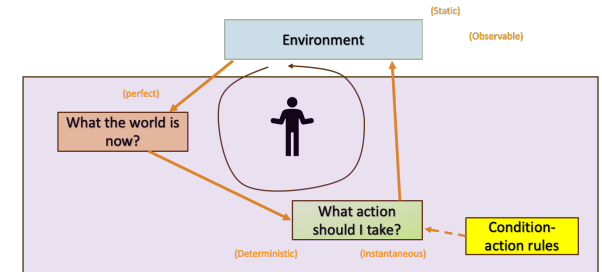
$$\frac{a \vee \beta, \quad \neg\beta}{a}$$

- ◇ **Resolution:** (This is the most difficult. Because 0 cannot be both true and false, one of the other disjuncts must be true in one of the premises. Or equivalently, implication is transitive.)

$$\frac{a \vee \beta, \quad \neg\beta \vee \gamma}{a \vee \gamma} \quad \text{or equivalently} \quad \frac{\neg\alpha \Rightarrow \beta, \quad \beta \Rightarrow \gamma}{\neg\alpha \Rightarrow \gamma}$$

Source: Russell & Norvig, AI: A Modern Approach

KB Agent Procedure



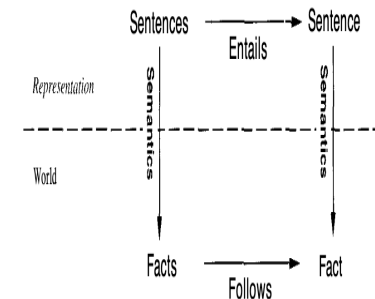
function KB-AGENT(*percept*) **returns** an *action*
static: *KB*, a knowledge base
t, a counter, initially 0, indicating time

TELL(*KB*, MAKE-PERCEPT-SENTENCE(*percept*, *t*))
action — ASK(*KB*, MAKE-ACTION-QUERY(*t*))
TELL(*KB*, MAKE-ACTION-SENTENCE(*action*, *t*))
 $t \leftarrow t + 1$
return *action*

Source: Russell & Norvig, AI: A Modern Approach

Formal Logic – 2/3

- Levels at which sentences are encoded
 - Epistemic (also called knowledge): what agents knows or believes
 - Logical: how sentences are encoded to allow inferencing. E.g., symbols
 - Executional: how sentences are encoded during execution. E.g., vectors, symbols
- Properties of sentences
 - **Valid**: A sentence is **valid** or necessarily true if and only if it is true under all possible interpretations in all possible worlds. Also called a **tautology**
 - **Satisfiable**: A sentence is satisfiable if and only if there is some interpretations in some possible worlds where it is true.



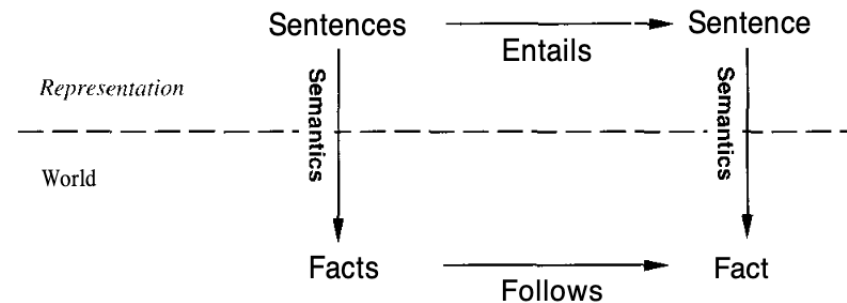
Credits:

- Russell & Norvig, AI - A Modern Approach
- Deepak Khemani - A First Course in AI

Formal Logic – 3/3

- Properties of Logic System

- **Soundness:** if it produces only true statements
- **Completeness:** if it produces all true statements
- **Consistency:** if it does not produce a sentence and its negation



Language	Ontological Commitment (What exists in the world)	Epistemological Commitment (What an agent believes about facts)
Propositional logic	facts	true/false/unknown
First-order logic	facts, objects, relations	true/false/unknown
Temporal logic	facts, objects, relations, times	true/false/unknown
Probability theory	facts	degree of belief 0...1
Fuzzy logic	degree of truth	degree of belief 0...1

Credits:

- Russell & Norvig, AI - A Modern Approach
- Deepak Khemani - A First Course in AI

Example: Course Selection

Example Situation – Course Selection

- A person wants to pass an academic program in two majors: A and B
- There are three subjects available: A, B and C, each with three levels (*1, *2, *3). There are thus 9 courses: A1, A2, A3, B1, B2, B3, C1, C2, C3
- To graduate, at least one course at beginner (*1) level is needed in major(s) of choice(s), and two courses at intermediate levels (*2) are needed
- **Answer questions**
 - Q1: How many minimum courses does the person have to take ?
 - Q2: Can a person graduate in 2 majors studying 3 courses only?
 - ...

Representation - Example

- Domain Description: “There are three subjects: A, B and C, each with three levels (*1, *2, *3).”
- Representation
 - has_studied_courseA1: yes – student has taken course; no – student has not taken
 - has_studied_courseA2
 - has_studied_courseA3
 - has_studied_courseB1
 - has_studied_courseB2
 - has_studied_courseB3
 - has_studied_courseC1
 - has_studied_courseC2
 - has_studied_courseC3

Issue: What about hierarchy among courses?

Representation - Example

- Domain Description: “There are three subjects: A, B and C, each with three levels (*1, *2, *3).”
- Representation
 - has_studied_courseA1: yes – student has taken course; no – student has not taken
 - has_studied_courseA2
 - has_studied_courseA3
 - has_studied_courseB1
 - has_studied_courseB2
 - has_studied_courseB3
 - has_studied_courseC1
 - has_studied_courseC2
 - has_studied_courseC3
- Previous statements set did not capture hierarchy between levels; new sentences would not have followed the reality in the world. Need more statements – LowerThan as shown.

```
LowerThan_Course_A1_CourseA2  
LowerThan_Course_A2_CourseA3  
LowerThan_Course_B1_CourseB2  
LowerThan_Course_B2_CourseB3  
LowerThan_Course_C1_CourseC2  
LowerThan_Course_AC_CourseC3
```

How to Tackle Course Selection Problem ?

- Represent the world as sentences in KB
 - Update KB based on scenarios
- Solve problems about courses selection scenarios
 - Pose problems as queries to KB
 - Interpret answers // reason with the world

```
function KB-AGENT(percept) returns an action  
static: KB, a knowledge base  
       t, a counter, initially 0, indicating time  
  
  TELL(KB, MAKE-PERCEPT-SENTENCE(percept, t))  
  action ← ASK(KB, MAKE-ACTION-QUERY(t))  
  TELL(KB, MAKE-ACTION-SENTENCE(action, t))  
  t ← t + 1  
  return action
```

Source: Russell & Norvig, AI: A Modern Approach

Major Types of Reasoning

- Inference: From premises to conclusions
 - Major types
 - **Deduction**: deriving logical conclusions from premises known or assumed to be true
 - **Induction**: deriving from particular premises to a universal conclusion.
 - **Abduction**: from an observation, find the most likely conclusion from the observations
- Usage
 - Deduction is useful to build knowledge bases from parts
 - Induction: to generalize
 - Abduction is a good source for hypothesis / priors in Bayesian learning

Setting Up for AIMA Code

- AI resources
<https://github.com/biplav-s/course-ai-tai-f23/blob/main/sample-code/AI-Resources.md>
- Setting up for Python code from AIMA book
<https://github.com/biplav-s/course-ai-tai-f23/tree/main/sample-code/ai-book-samples>

Exercise and Code

- Logical Reasoning

- From Book: AI – A Modern Approach, <https://github.com/aimacode/aima-python/blob/master/logic.ipynb>

1,4	2,4	3,4	4,4
1,3 W!	2,3	3,3	4,3
1,2 A S OK	2,2 OK	3,2	4,2
1,1 V OK	2,1 B V OK	3,1 P!	4,1

A = Agent
B = Breeze
G = Glitter, Gold
OK = Safe square
P = Pit
S = Stench
V = Visited
W = Wumpus

Source: Russell & Norvig, AI: A Modern Approach

Examples of Agents

Agent Type	Percepts	Actions	Goals	Environment
Medical diagnosis system	Symptoms, findings, patient's answers	Questions, tests, treatments	Healthy patient, minimize costs	Patient, hospital
Satellite image analysis system	Pixels of varying intensity, color	Print a categorization of scene	Correct categorization	Images from orbiting satellite
Part-picking robot	Pixels of varying intensity	Pick up parts and sort into bins	Place parts in correct bins	Conveyor belt with parts
Refinery controller	Temperature, pressure readings	Open, close valves; adjust temperature	Maximize purity, yield, safety	Refinery
Interactive English tutor	Typed words	Print exercises, suggestions, corrections	Maximize student's score on test	Set of students

Source: Russell & Norvig, AI: A Modern Approach

Lecture 12: Summary

- We talked about
 - Knowledge-based agents
 - Logic (Propositional)
 - Inferencing (Propositional)

Week 6: Concluding Comments

We talked about

- AI/ ML Trust
 - Explainability
 - Trust ratings
- Representation and Logic
 - Propositional

- Week 1: Introduction, Aim: Chatbot / Intelligence Agent
- Weeks 2-3: Data: Formats, Representation and the Trust Problem
- Week 3: Machine Learning – Supervised (Classification)
- Week 4: Machine Learning - Unsupervised (Clustering) –
- Topic 5: Learning neural network, deep learning, Adversarial attacks
- Week 6: Large Language Models – Representation and Usage issues
- Weeks 7-8: Search, Heuristics - Decision Making
- Week 9: Constraints, Optimization – Decision Making
- Topic 10: Markov Decision Processes, Hidden Markov models -
Decision making
- Topic 11-12: Planning, Reinforcement Learning – Sequential decision making
- Week 13: Trustworthy Decision Making: Explanation, AI testing
- Week 14: AI for Real World: Tools, Emerging Standards and Laws; Safe AI/ Chatbots

Upcoming Evaluation Milestones

- Projects B: Sep 30 – Nov 20
- Quiz 2: Oct 7
- Quiz 3: Nov 11
- Paper presentation (grad students only) : Nov 18

About Week 7 – Lectures 13 and 14

Week 7 – Lectures 13 and 14

- L13: Logic and Inference - First Order
- L12: Search, Search - Uninformed

- Week 1: Introduction, Aim: Chatbot / Intelligence Agent
- Weeks 2: Data: Formats, Representation, ML Basics
- Week 3: Machine Learning – Supervised (Classification)
- Week 4: Machine Learning - Unsupervised (Clustering) –
- Topic 5: Learning neural network, deep learning, Adversarial attacks
- Week 6: Large Language Models – Representation and Usage issues
- Weeks 7-8: Search, Heuristics - Decision Making
- Week 9: Constraints, Optimization – Decision Making
- Topic 10: Markov Decision Processes, Hidden Markov models -
Decision making
- Topic 11-12: Planning, Reinforcement Learning – Sequential decision making
- Week 13: Trustworthy Decision Making: Explanation, AI testing
- Week 14: AI for Real World: Tools, Emerging Standards and Laws; Safe AI/ Chatbots

Note: exact schedule changes slightly to accommodate for exams and holidays.