Introduction

Professor Marie Roch

Chapters 1 & 2, 27.3, Russell & Norvig

What is AI?

Riken and Sumitomo nursing robot

“Machines with human skills”

- Michelle Zhou, CEO Juji Inc.

Intelligence per Merriam Webster

1 a : 1) the ability to learn or understand or to deal with new or trying situations **:** REASON; *also* **:** the skilled use of reason

2) the ability to apply knowledge to manipulate one's environment or to think abstractly as measured by objective criteria (as tests)

b *Christian Science* **:** the basic eternal quality of divine Mind

c : mental acuteness **:** SHREWDNESS

2 a : an intelligent entity; *especially* **:** ANGEL

b : intelligent minds or mind <cosmic *intelligence*>

3 : the act of understanding **:** COMPREHENSION

Artificial intelligence

- weak AI - Machines act as if they are intelligent

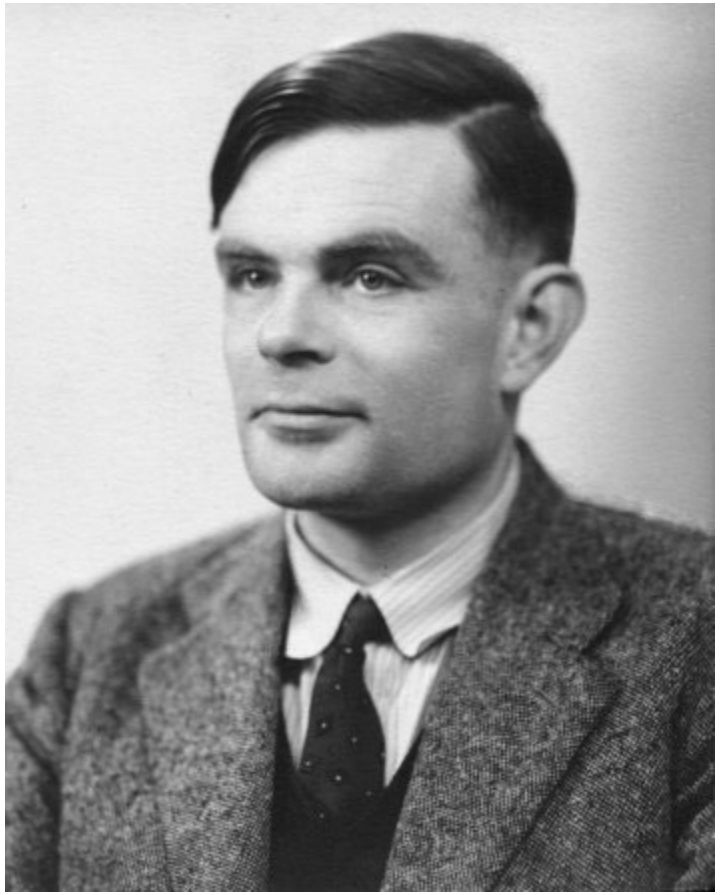
- strong AI – Machines think

Brent Spiner as Commander Data
Star Trek Next Generation

What does it mean for a machine to be intelligent?

Turing Test – Can a human tell that they are interacting with a computer?

Variant of this competition occurs today with Loebner Prize (very restricted Turing test for chatbots)

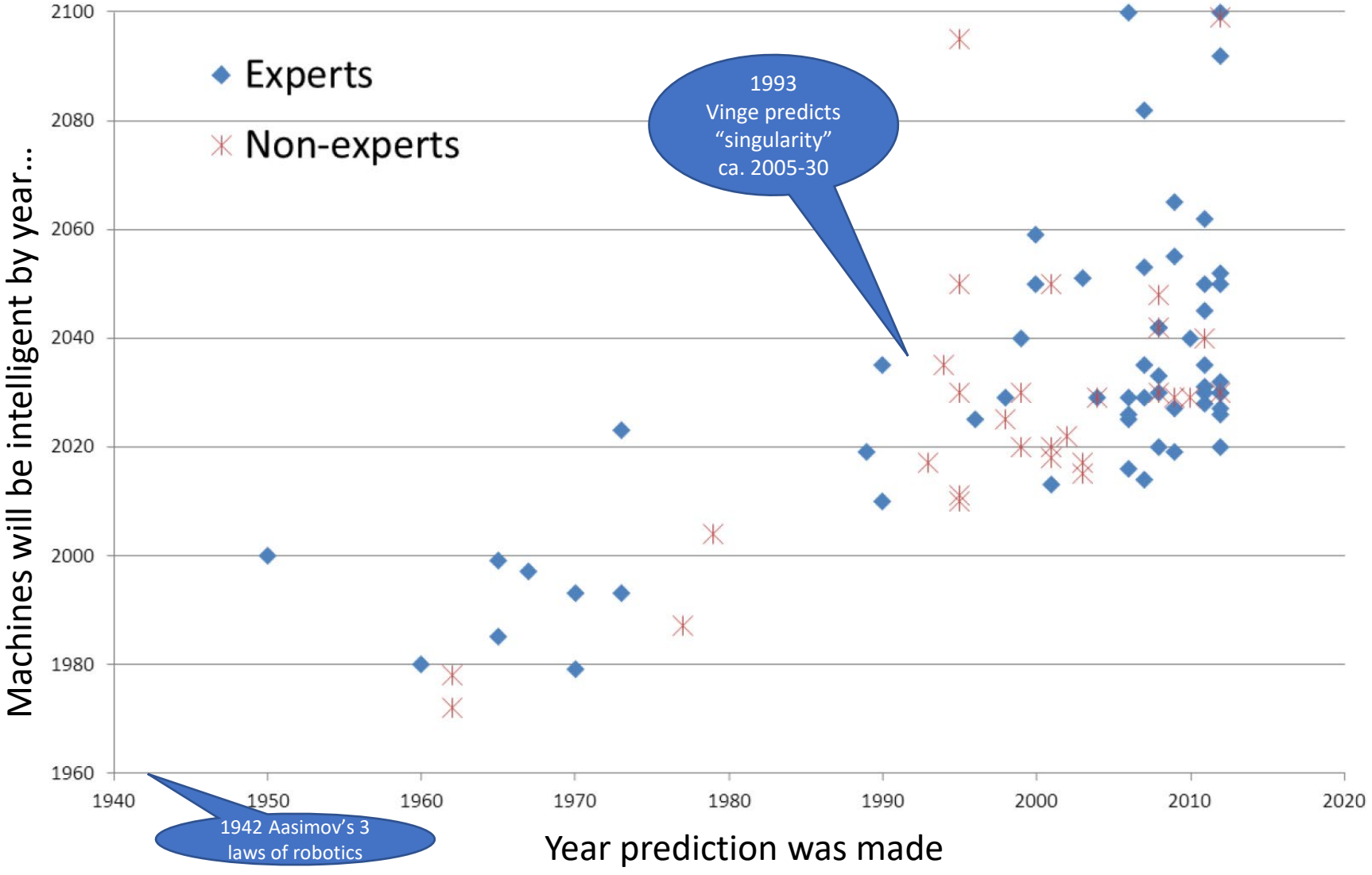


Alan Turing 1912-1954
See *The Imitation Game* (2014) for a dramatization of his life.

Artificial intelligence

- Turing test - acting intelligent is sufficient
- Cognitive approach – Model the thinking of humans or other animals
- Laws of thought – Set of logic rules
  drawback: difficult to model informal/general knowledge
- Rational agent – Logical inference is good, but not always applicable.

When will we have strong AI? (Armstrong, et al. 2014)



Eliza (Weizenbaum, 1966)

young woman: Men are all alike.
eliza: In what way?
young woman: They're always bugging us about something specific or other.
eliza: Can you think of a specific example?
young woman: Well, my boyfriend made me come here.
eliza: Your boyfriend made you come here?
young woman: He says I'm depressed much of the time.
eliza: I am sorry to hear you are depressed.
young woman: It's true. I'm unhappy.
eliza: Do you think coming here will help you not to be unhappy?
young woman: I need some help, that much seems certain.
eliza: What would it mean to you if you got some help?
young woman: Perhaps I could learn to get along with my mother.

Simple rules ---> disappointment

(X me Y) ---> (X you Y)

(I remember X) ---> (Why do you remember X just now?)

(My {family-member} is Y) ---> (Who else in your family is Y?)

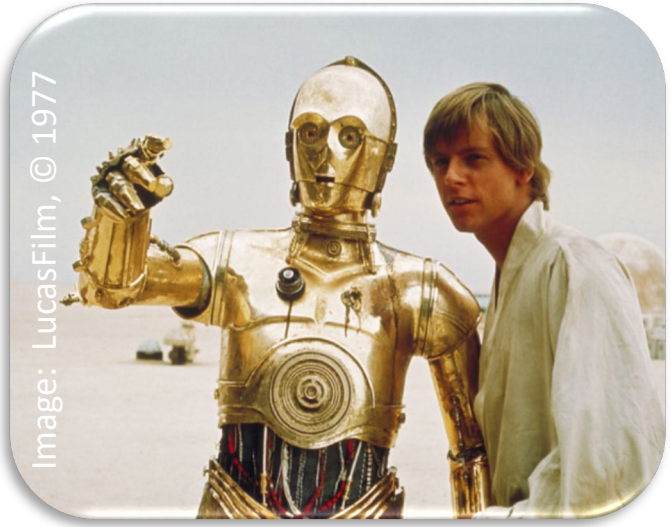
(X {family-member} Y) ---> (Tell me more about your family)

Areas contributing to AI

- Philosophy
- Mathematics
- Neuroscience
- Psychology
- Linguistics
- Computer science

- Many application areas contribute as well (e.g. economics)

Intelligent agents



- Sensors provide perceptual input (**percepts**) of environment

- Agent makes decisions
- Actions carried out through actuators that may affect the environment

Intelligent agents

- Softbot – Software only agent
  - Available data are percepts
  - Examples: web-based reputation monitoring, game opponent

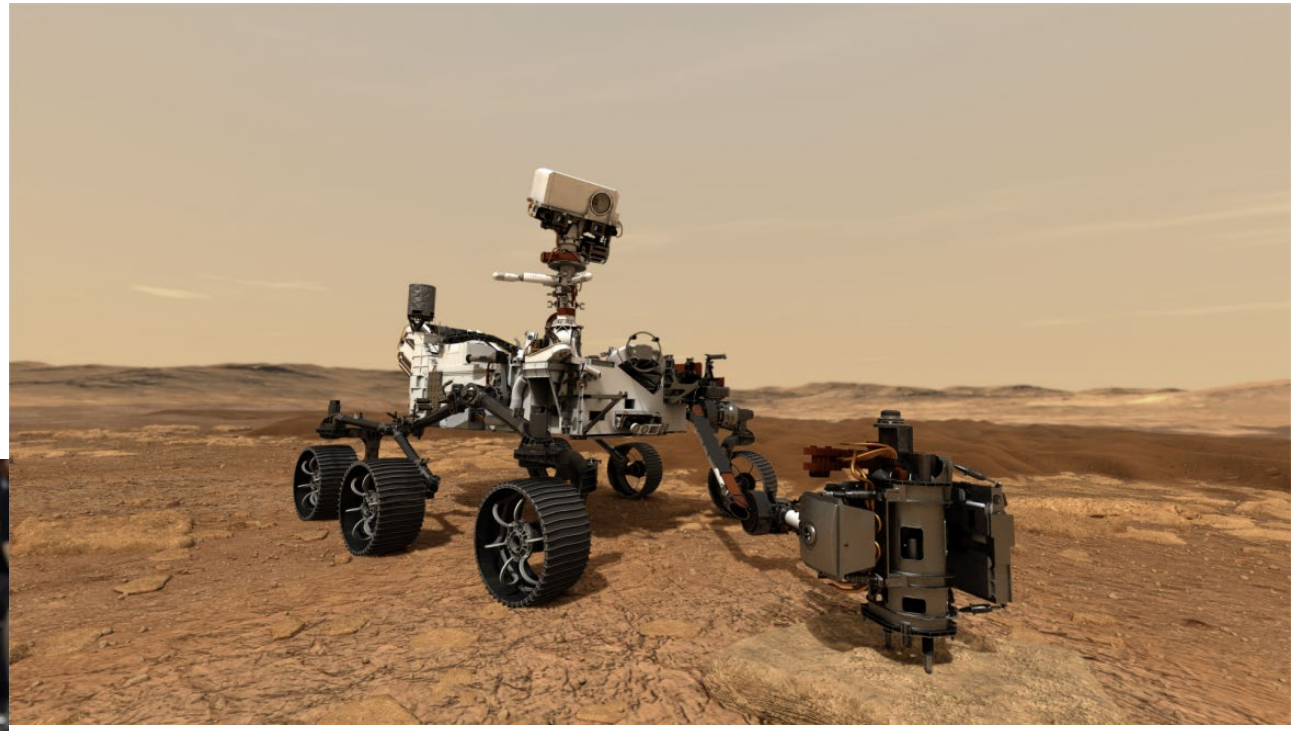
Task environments

In what environment will the agent be operating?

- fully vs. partially observable
  - partially observable → uncertain state
- Rules are
  - known: Outcome (or outcome probabilities) are known
  - unknown – Outcomes must be learned

Task environments

Getty images

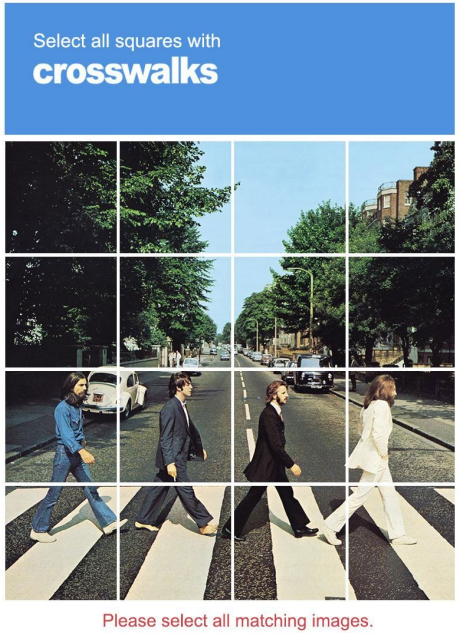


NASA Perseverance

- single- vs. multi- agent
- multiagent issues
  - cooperative vs. competitive
  - communication
  - randomization to prevent predictability

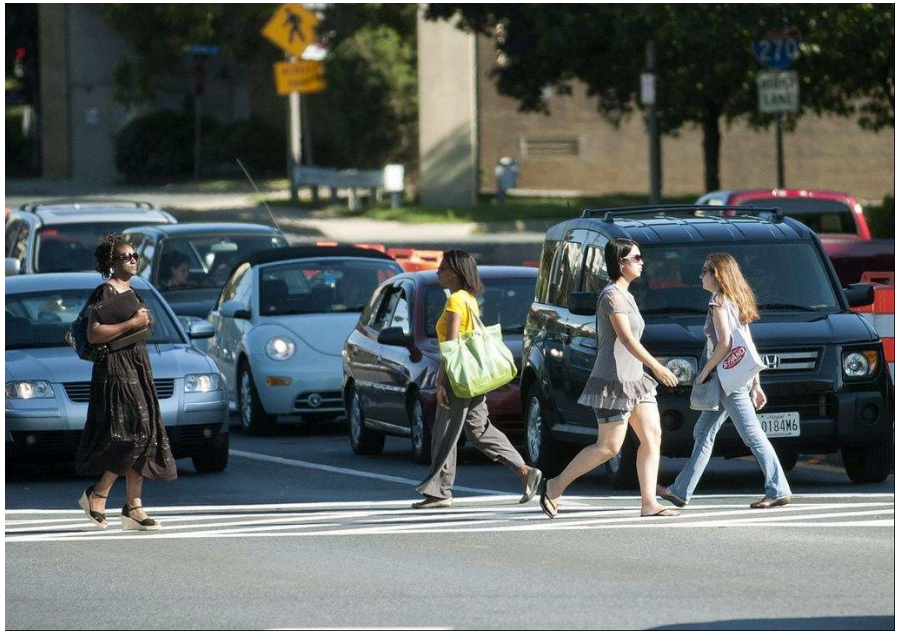
Task environments

- What happens when an agent acts?
  - deterministic – we know next state
  - stochastic
    - nondeterministic factors may influence (stochastic →probabilities) leading to an *uncertain* state
- Decisions are
  - episodic – Next decision only depends on state
  - sequential – Next decision dependent on previous ones and can affect future states



Task environments

- State can be
  - static – does not change while agent is deciding next action
  - dynamic – Environment constantly changing
  - semidynamic – Environment static, but performance is time dependent



See Figure 2.6 (p. 47) for example task environments

Agent structure

- An agent's architecture consists of
  - data structures
  - code
- Simplest agent: table driven

```
function table-driven-agent(percept) returns action
  persistent: percepts (sequence, empty at first)
              table of actions indexed by
                    percept sequence


  percepts.append(percept)
  return lookup(percepts, table)
```

What's wrong here?

Agent types

- Simple reflex – Reacts to stimulus
- Model-based reflex – Stimulus + state
- Goal-based – Work towards objective
- Utility-based – Increase value of something measurable
- Learning – Adjust goals/utility/rules over time

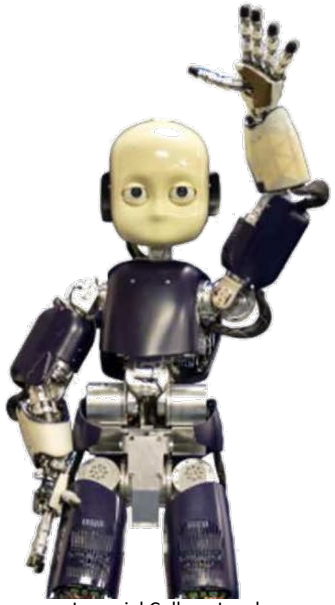
Simple-reflex agents

- Ignores percept history, uses the current one
- Productions (aka conditions-action) decide action, e.g.

  person waving → wave

  person smiling → smile

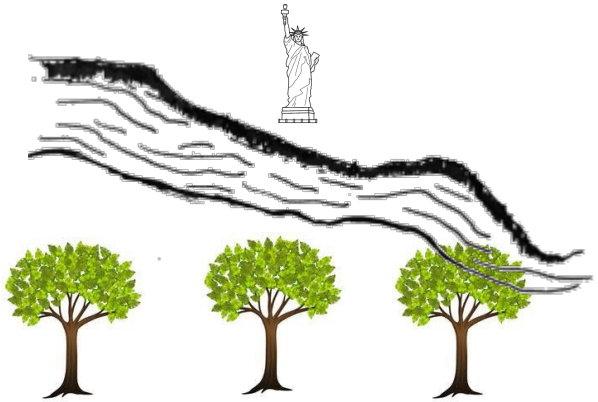
  person swinging hammer towards me → duck!



source: Imperial College London

Model-based reflex agents

- Add internal state
- New percepts update the state
- Productions based on percept and state



no credit available

The Nature Conservancy

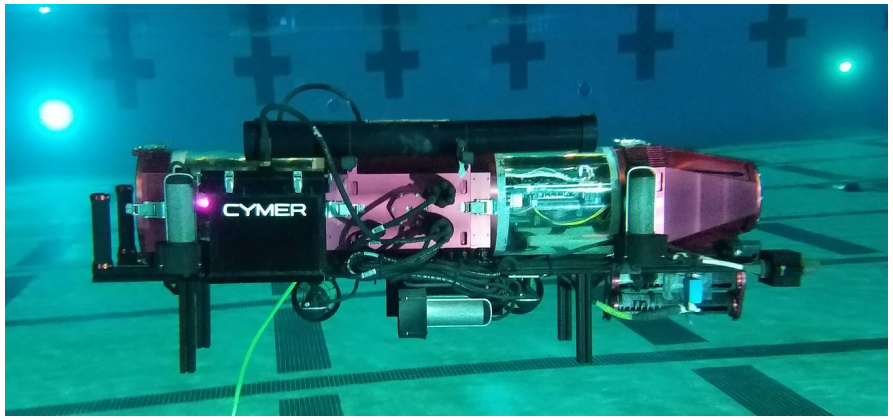
States in model-based agents

States can be

- atomic – state is indivisible, it does not have multiple parts
  sometimes we treat things as atomic, even if they are not: e.g. configuration of pieces on a board

- factored – multiple attributes
  example: autonomous underwater vehicles have representations for pressure, power, buoyancy, sonar, vision, …

SAN DIEGO STATE UNIVERSITY MECHATRONICS



SDSU Mechatronics Robosub AUV

Goal-based agents

- Agent works to achieve a specific state
- Usually requires: Search and Planning

Utility-based

- Based on utility theory: The idea of how useful or happy something makes you.

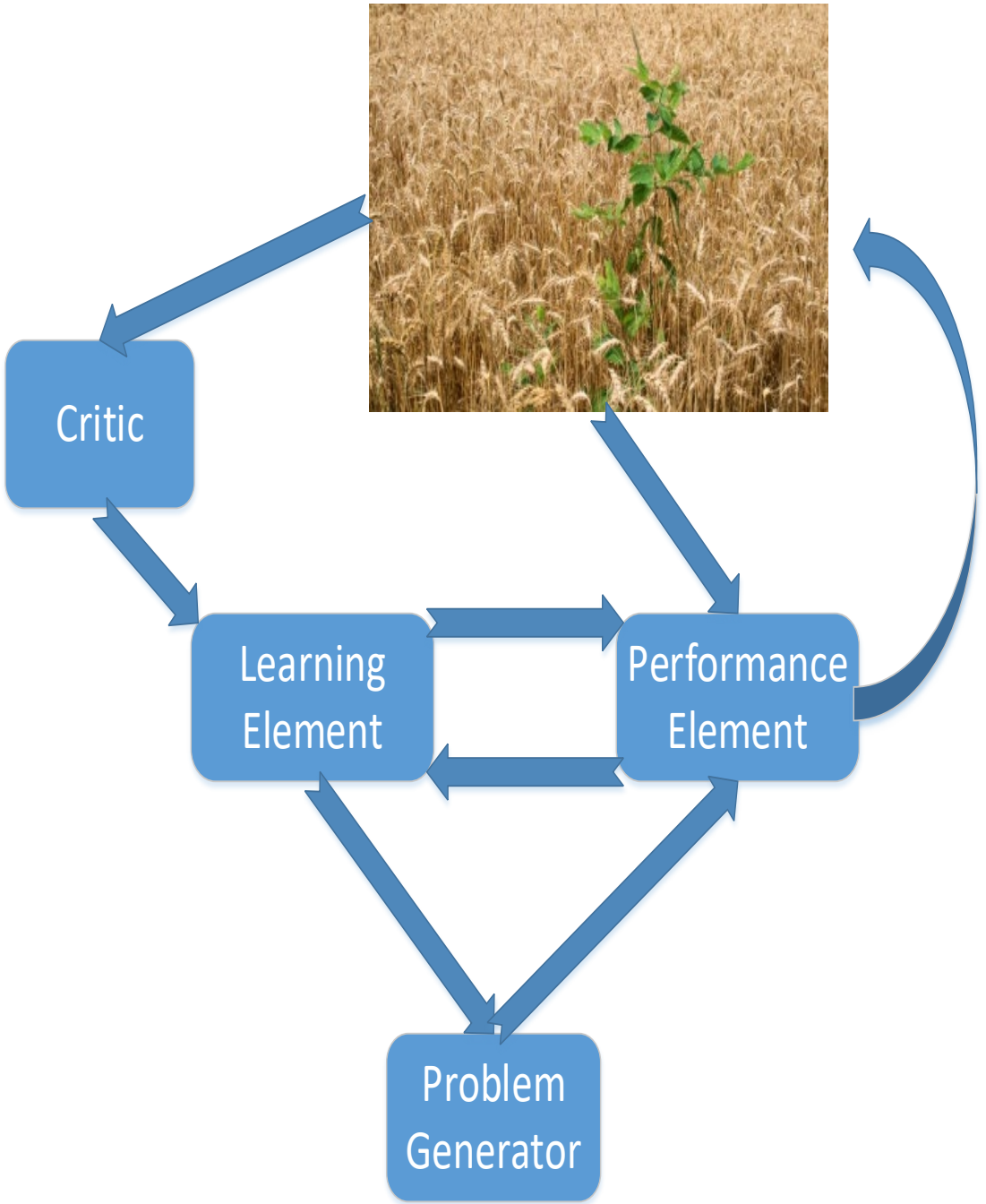
15 utils

-5 utils

- Decisions are made to maximize the expected utility.

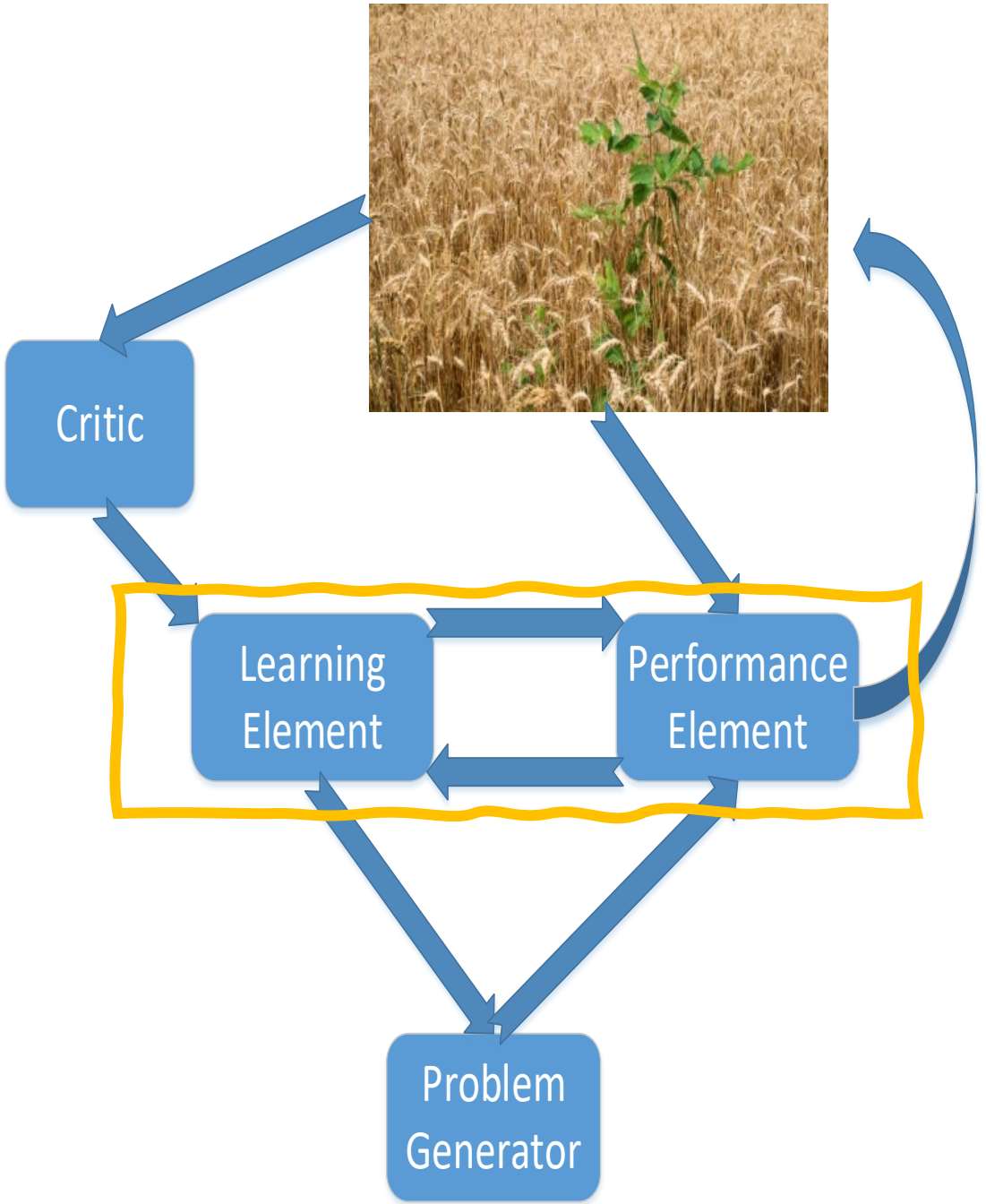
Learning agents

Decision rules are adjusted based on performance over time



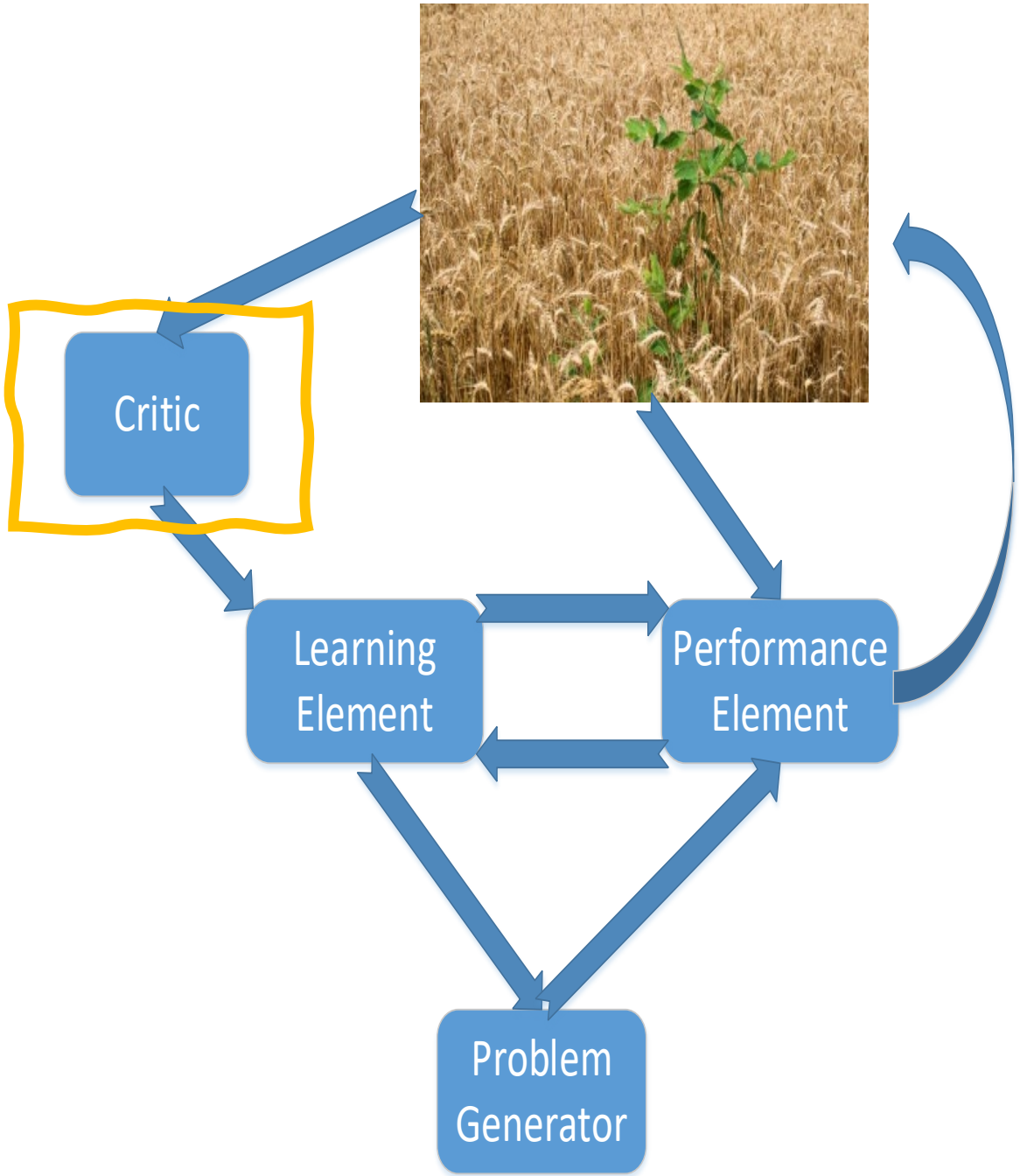
Learning agents

- Performance element makes decisions based on percepts
- Learning element uses knowledge from performance element to modify rules



Learning agents

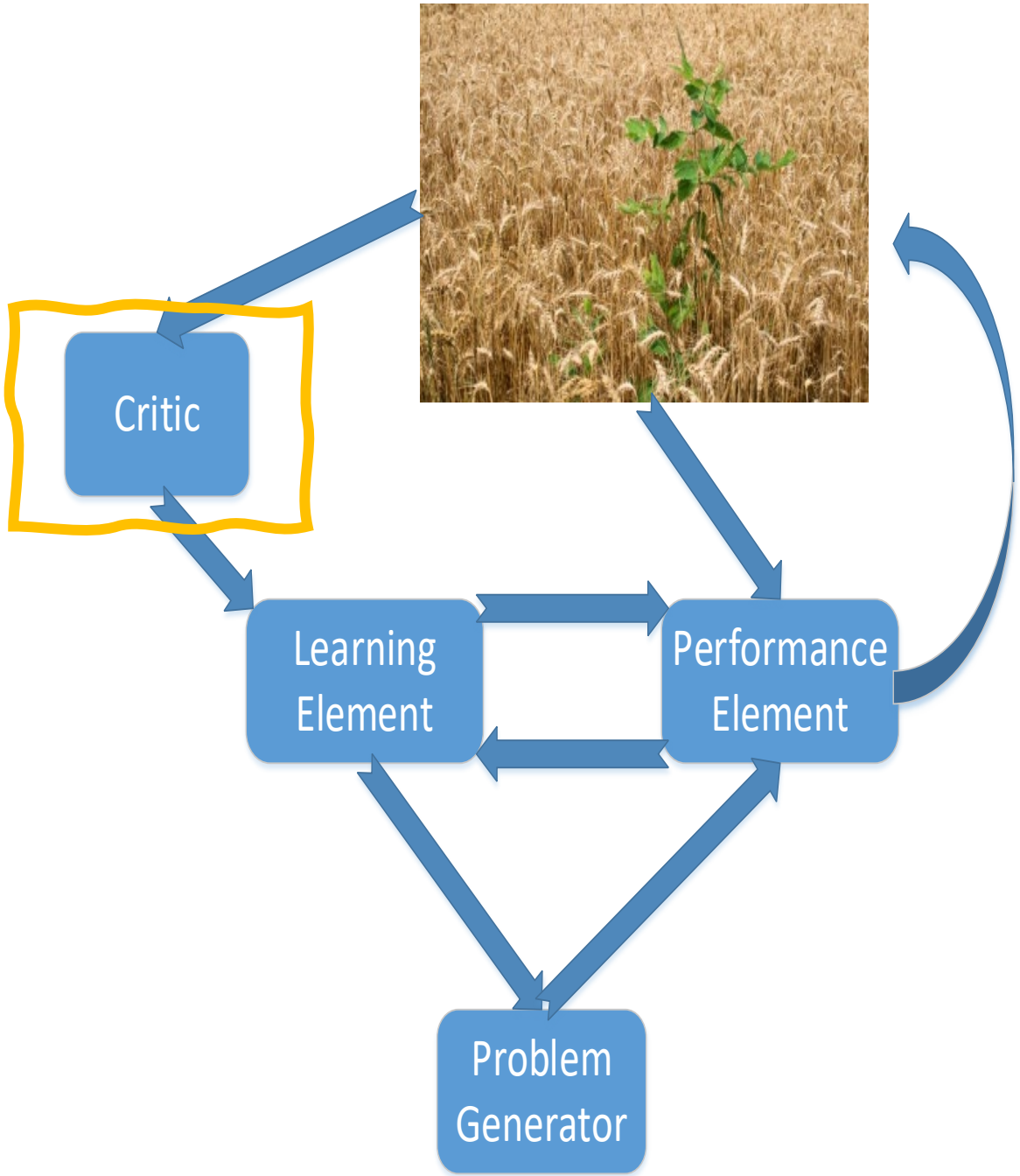
Critic monitors the environment to see how the system is doing and keeps learning element from going off track.



Learning agents

Critic monitors the environment:

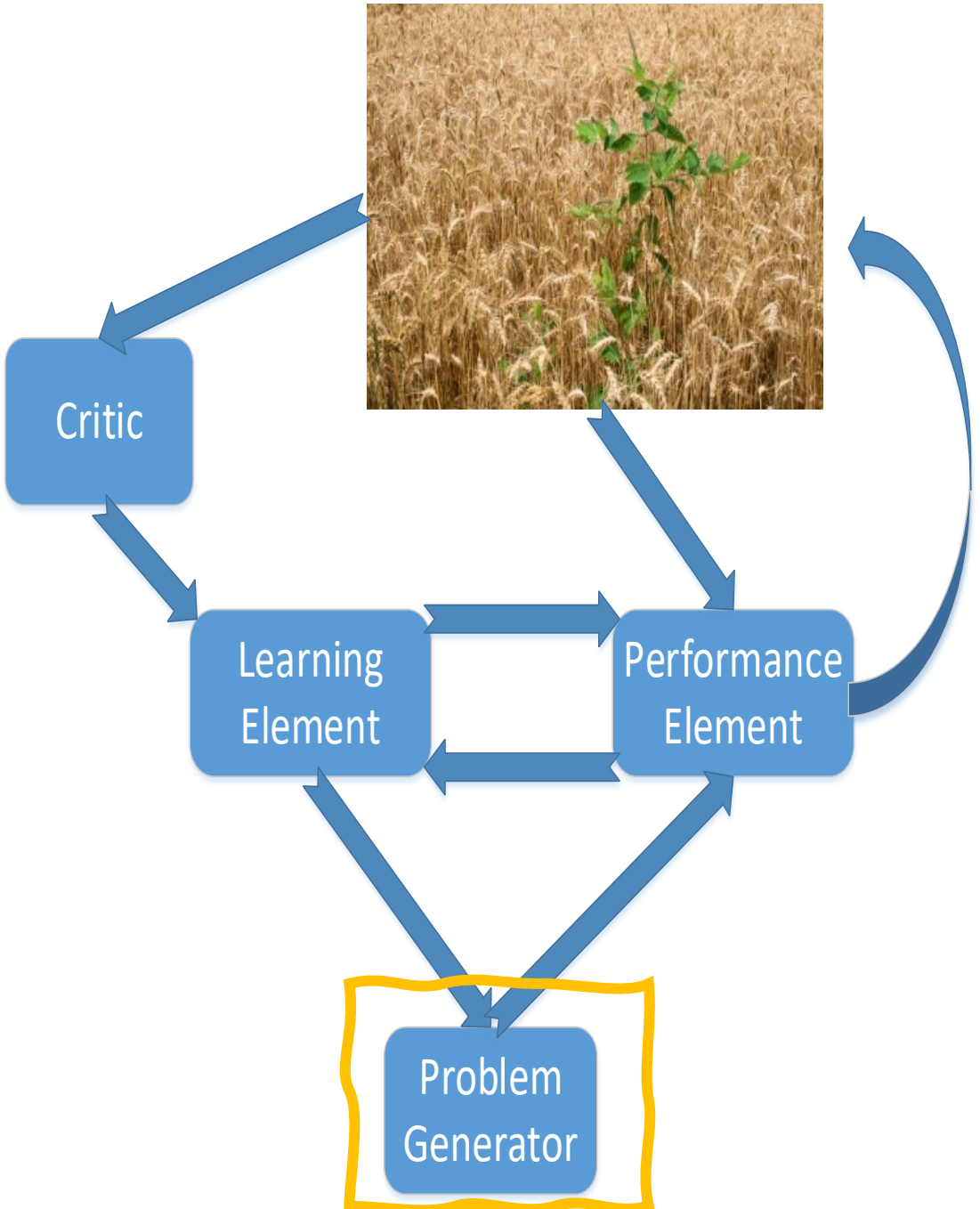
- Provides utility of the current percepts according to an unchanging standard
- Allows learning element to learn useful goals



Learning agents

Problem generator

- Uses learning element's goals to suggest experiments
- Experiments may lead to the learning element improving the performance element



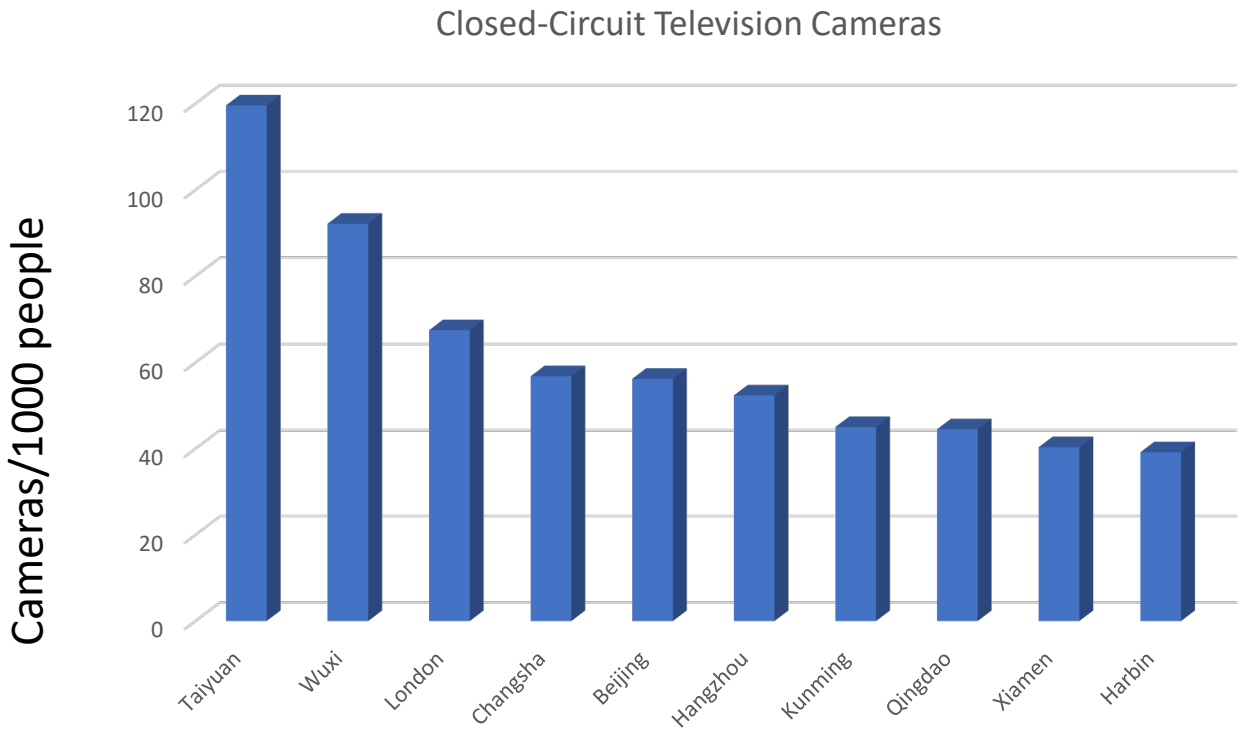
Ethical issues

- Autonomous weapon systems
  - Is it ethical to have lethal systems without a human in the loop?
  - Issues:
    - Should machines be making decisions that result in loss of life?
    - What about errors?
  - Example: Harop missile (Israel) – Can be configured to autonomously recognize targets and launch.

wikipedia.org/wiki/IAI_Harop

Ethical issues

- Surveillance, security, privacy



Data Source:
Keegan, Matthew, "The Most Surveilled Cities in the World", *US News and World Report*, August 14, 2020.

Ethical issues

- Surveillance – Cameras, microphones, online activity can be coupled with AI systems to recognize, track, and flag people/behaviors
  - Government and corporate users
  - Misidentification can create large problems for individuals

Thumbnails from Cao et al's Open Pose (2019)
full video

Ethical issues

- Security/Privacy
  - As data are amassed, misuse can occur from
    - corporate/government actions
    - data breaches
  - Protection strategies
    - De-identification – Remove data that can be linked to individuals
    - Problem: Remaining data can be traced to individuals
      Sweeney showed that 87% of Americans can be uniquely identified from
      - date of birth
      - gender
      - zip code
  - People frequently do not know what data have been amassed and how it is used.

Ethical issues

- Fairness and bias
  - Machine learning algorithms learn from the data they are given
  - Example: face recognition (Manyika, Silberg, Presten (2019) *Harvard Business Review*)
    - Joy Buolamwini found higher errors on minority women (and minorities in general).
    - Potential root cause: lack of representation in training data

Joy Buolamwini MIT News

Ethical issues

- Trust and transparency
  - How good are our systems?
    - Are they robust?

Being chauffeured by a Tesla
(don't do this)

deepdrive.berkeley.edu: Stop signs recognized as 45 MPH speed limit, right turn as stop

  - What have we done to validate that they work?
  - Are decisions *explainable*?

Semester outline

- Search
- Machine learning
- Vision
- Constraint satisfaction
- Logic
- Uncertainty
- Natural language