Artificial Intelligence
1 – Introduction

Arthur Bit-Monnot

INSA 4IR

Section 1

What is AI?

What is AI?

- AI is a field that aims at *building* intelligent entities.

What is intelligence?

Acting humanly: The Turing Test

Alan Turing (1950): "Can machines think?"

A human interrogator interacts with a human and a machine, writing questions and receiving written answers.

If the interrogator cannot distinguish between the two, the machine is intelligent.

Acting humanly: The Turing Test

Alan Turing (1950): "Can machines think?"

A human interrogator interacts with a human and a machine, writing questions and receiving written answers.

If the interrogator cannot distinguish between the two, the machine is intelligent.

Needed capabilities:

- Natural language processing: understanding and generating text.
- Knowledge representation: storing and using information.
- Automated reasoning: answering questions.
- Machine learning: adapting to new circumstances.

Acting humanly: The Turing Test

$\rightarrow$ ethnocentric: we consider ourselves as the most intelligent beings, therefore an intelligent machine should be able to mimic us.

Application to aeronautics:

- the peregrine falcon is the best flyer (diving speed of 390km/h!!!)
- the *peregrine test*:

> *A machine is capable of flying if it can fly so perfectly that it is indistinguishable from a peregrine falcon.*

Acting rationally: The rational agent

- An **agent** is an entity that **perceives** its environment and **acts** upon it.
- An agent is **rational** if it acts so as to achieve the best (expected) outcome

Acting rationally: The rational agent

- An **agent** is an entity that **perceives** its environment and **acts** upon it.
- An agent is **rational** if it acts so as to achieve the best (expected) outcome

- Rationality is independent of:
  - the thought process
  - human behavior

*AI focus: study and constructions of agents that **do the right thing**.*

A Go Rational Playing Agent

- perceives the current state of the board
- select the move that maximizes the probability of winning

A Go Rational Playing Agent

- perceives the current state of the board
- select the move that maximizes the probability of winning

Often, finding the *best* move is impossible due to computational limits.

> *Limited rationality: find a good move in a reasonable time and under computational limits.*

Rational Agent Model

- s: the current state of the environment (a situation)
- $\mathcal{A}(s)$: the set of actions available in state s
- $utility(s, a)$: the expected utility of taking action a in state s

$$act(s) = \arg\max_{a \in \mathcal{A}(s)} utility(s, a)$$

- Given: objective function (utility)
- AI: implement `argmax`: search for the best action

Objective Functions and Beneficial Machines

*The objective function is **the** key to the behavior of the agent.*

- Chess/Go: win the game
- Self-driving car: ???
  - progress towards destination
  - ensure safety
  - minimize fuel consumption
  - ...

Objective Functions and Beneficial Machines

*The objective function is **the** key to the behavior of the agent.*

- Chess/Go: win the game
- Self-driving car: ???
  - progress towards destination
  - ensure safety
  - minimize fuel consumption
  - ...

- How should we balance these objectives to align it with our own objective?
  - value alignment problem (now and in the future)

Misbehaving Intelligent Agents

I replace my chess-solving software by a fully capable robot (so that I don't have to move the pieces myself).

What could go wrong?

Misbehaving Intelligent Agents

I replace my chess-solving software by a fully capable robot (so that I don't have to move the pieces myself).

What could go wrong?

- With the sole objective of winning:
  - the robot could cheat (e.g., by moving the pieces when I'm not looking)
  - the robot could blackmail its opponent
  - the robot could drug its opponent

Not "unintelligent" behavior, but *misaligned* with our (implicit) objectives.

Section 2

AI: Current State, Risks and Opportunities

"Intelligent Systems": State of the Art

- **Automated translation**: support for 100+ languages with close to human performance in most represented languages.
- **Game playing**: Outperforming humans in chess, Go, poker, etc.
- **Legged locomotion**: quadruped robots can walk, run, jump, and even do backflips.
- **Planning and Scheduling**: Optimal scheduling of tasks in complex environments over long horizons.
- **Image understanding**: Recognizing objects, people, actions, scenes in images and videos.
- ...

General and Super Artificial Intelligence

Currently, many domains where *specialized* AI systems outperform humans.

- **General AI**: AI that can perform any intellectual task that a human can do.
- **Artificial Superintelligence**: AI that vastly surpasses human intelligence in every field.

No consensus on the feasibility and time frame of these developments.

- high (existential!) risks
- high opportunities: solving complex problems, improving human life

> *Value alignment problem: how to ensure that a superintelligent AI's objectives are aligned with ours?* [1]

[1] Nick Bostrom, Superintelligence: Paths, Dangers, Strategies, 2014

In the very short term

Fundamentally, the field of AI is about building intelligent agents, that may do tasks that were previously only doable by humans.

$\Rightarrow$ Replaces requirement on man-power (which requires money and persuasion!) by the much more affordable machine-power.

In the very short term

Fundamentally, the field of AI is about building intelligent agents, that may do tasks that were previously only doable by humans.

$\Rightarrow$ Replaces requirement on man-power (which requires money and persuasion!) by the much more affordable machine-power.

Opportunities:

- **Productivity**: many complex tasks do not require costly human intervention

Risks:

- **Lethal autonomous weapons**: requiring no human supervision makes them very scalable
- **Surveillance**: tracking individuals, predicting behavior
- **Persuasion**: individual targeted propaganda, etc.

AI and Society

Additional challenges:

- *accountablity*: who is responsible for the actions of an AI?
- *transparency*: how to understand the decisions of an AI?
- *fairness*: how to ensure that the AI does not discriminate?

Other opportunities:

- Finding much better solutions to complex problems (e.g., railway optimization)
- Cracking open problems that were previously intractable (e.g., protein folding)

Section 3

Structure of the course

Objectives and Content

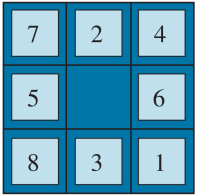
- Define AI and its scope
- Understand the basic concepts of AI
- Structure of intelligent agents
- Classification of AI problems/environments
- **Sequential decision making**: search, planning, (reinforcement learning)

Related courses

- Reasoning and Knowledge Representation: logic (4IR), Ontologies (4IR)
- Optimization: linear optimization (3MIC), metaheuristics (4IR)
- Problem solving: Constraint Programming / SAT (5SDBD)
- Machine learning: supervised (4IR), unsupervised (5A)

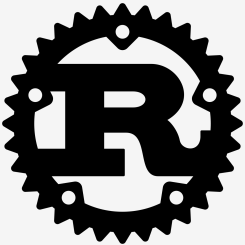
Labs: Sequential decision making

- Puzzles: Single player, deterministic (1 lab)
  - 8-puzzle (fr: taquin) → best-first search
- Single Player Games (non-deterministic) (1 lab)
  - 2048 → expectimax, Monte-Carlo Tree Search (MCTS)
- Multi-Player games (2 labs)
  - Risk-like (Monte Carlo Tree Search, reinforcement learning?)

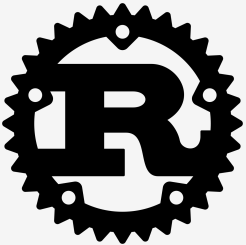


Labs: practical details

- All labs will be done in **Rust**
- why?
  - performance (runtime & startup)
  - reliability



Labs: practical details



- All labs will be done in **Rust**
- why?
  - performance (runtime & startup)
  - reliability

Rust has a steep learning curve:

- will only require the fundamentals
- will provide a lot of help and examples

But you should get familiar with the language ahead of the labs

- Website: `Labs/Rust`
- reference: The Rust Programming Language (online book)

Evaluation

- 1 written exam (~50%)
- project-based evaluation (~50%)
  - based on the last 2 labs
  - precise modalities to be determined
  - most likely: concise report + oral

Reference: AIMA

Artificial Intelligence: A Modern Approach (AIMA),
Stuart Russell and Peter Norvig
Fourth Edition, 2020

- Course will mostly follow the book structure.
- Most figures are from the book.