Artificial Intelligence Foundation – JC3001

Lecture 2: Agents I

Prof. Aladdin Ayesh (aladdin.ayesh@abdn.ac.uk)

Dr. Binod Bhattarai (binod.bhattarai@abdn.ac.uk)

Dr. Gideon Ogunniye, (g.ogunniye@abdn.ac.uk)

September 2025

UNIVERSITY OF ABERDEEN

Material adapted from:
Russell and Norvig (AIMA Book): Chapter 2
Michael Wooldridge (An Introduction to Multi-Agent Systems)

- Part 1: Introduction
  1. Introduction to AI ✓
  2. **Agents**
- Part 2: Problem-solving
  1. Search 1: Uninformed Search
  2. Search 2: Heuristic Search
  3. Search 3: Local Search
  4. Search 4: Adversarial Search
- Part 3: Reasoning and Uncertainty
  1. Reasoning 1: Constraint Satisfaction
  2. Reasoning 2: Logic and Inference
  3. Probabilistic Reasoning 1: BNs
  4. Probabilistic Reasoning 2: HMMs

- Part 4: Planning
  1. Planning 1: Intro and Formalism
  2. Planning 2: Algos and Heuristics
  3. Planning 3: Hierarchical Planning
  4. Planning 4: Stochastic Planning
- Part 5: Learning
  1. Learning 1: Intro to ML
  2. Learning 2: Regression
  3. Learning 3: Neural Networks
  4. Learning 4: Reinforcement Learning
- Part 6: Conclusion
  1. Ethical Issues in AI
  2. Conclusions and Discussion

- Motivate the abstraction of agents
- Introduce the notion of intelligent agent

▶ Background on Agents

▶ Agents

▶ Intelligent Agents

- Ubiquity
- Interconnection
- Intelligence
- Delegation
- Human-orientation

- Continuous reduction in cost of computing
- Processing power in contexts/devices which once were uneconomic
- Processing capability increases and spreads
- Sophistication and "intelligence" become <u>ubiquitous</u> (i.e., everywhere)
- £1M question: what could benefit from a processor?

- Computer systems no longer stand-alone
- System-of-systems networked as very large distributed systems
- Distributed and concurrent systems have become the norm
  — Researchers and practitioners developing (theoretical) models of computing as interaction among many distributed systems

- The complexity of tasks that we are capable of automating and delegating to computers has grown steadily
- If you don't feel comfortable with this definition of "intelligence", it's probably because you are a human

- Computers are doing more for us . . . without our intervention
- We are **giving control** to computers, even in safety critical tasks
- Example: fly-by-wire aircraft, where the machine's judgement may be more trusted than an experienced pilot
- Ongoing: fly-by-wire cars, intelligent braking systems, cruise control that maintains distance from car in front. . .

- Computing moving away from machine-oriented views of programming
- Heading toward concepts and metaphors that more closely reflect the way humans (us!) understand the world
- Programmers (and users!) relate to computers differently
- Programmers conceptualise and implement software in terms of ever higher-level – more <u>human-oriented</u> – abstractions

- Programming has progressed through:
    - Machine code;
    - Assembly language;
    - Machine-independent programming languages;
    - Sub-routines;
    - Procedures and function;
    - Abstract data types;
    - Objects;
        ...to...
    - Agents

- What techniques might be needed to deal with systems composed of 10^{10} processors?
- Don't be deterred by its seeming to be "science fiction"
- Hundreds of millions of people connected by email once seemed to be "science fiction" ...
- Assume that current software development models cannot handle this...

- Delegation and Intelligence imply the need to build computer systems that can act effectively on our behalf
- This implies:
  — The ability of computer systems to act independently
  — The ability of computer systems to act in a way that represents our best interests while interacting with other humans or systems

- Interconnection and Distribution have become core motifs in Computer Science
- But Interconnection and Distribution, coupled with the need for systems to represent our best interests, implies systems that can **cooperate** and **reach agreements** (or even **compete**) with other systems that have different interests (much as we do with other people)

- All of these trends have led to the emergence of a new field in Computer Science: **multiagent systems**

Here is one possible definition:

> *An agent is a computer system that is capable of **independent** action on behalf of its user or owner (figuring out what needs to be done to satisfy design objectives, rather than constantly being told)*

As for Multi-Agent Systems (MAS):

- A multiagent system is one that consists of a number of agents, which **interact** with one-another
- In the most general case, agents will be acting on behalf of users with different goals and motivations
- To successfully interact, they will require the ability to **cooperate**, **coordinate**, and **negotiate** with each other

- Two key problems in Multi-Agent Systems:
  - How do we build agents capable of independent, autonomous action, so that they can successfully carry out tasks we delegate to them? **(Agent Design)**
  - How do we build agents that are capable of interacting (cooperating, coordinating, negotiating) with other agents in order to successfully carry out those delegated tasks, especially when the other agents cannot be assumed to share the same interests/goals? **(Society Design)**

In Multi-Agent Systems, we address questions such as:

- How can cooperation emerge in societies of self-interested agents?
- What kinds of languages can agents use to communicate?
- How can self-interested agents recognise conflict, and how can they (nevertheless) reach agreement?
- How can autonomous agents coordinate their activities so as to cooperatively achieve goals?

While these questions are all addressed in part by other disciplines (notably economics and social sciences), what makes the multiagent systems field unique is that it emphasises that the agents in question are **computational**, **information processing** entities.
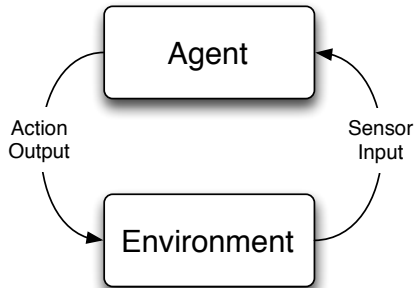
▶ Background on Agents

▶ Agents

▶ Intelligent Agents

- The main point about agents is they are *to some degree* **autonomous**: capable of acting independently, exhibiting control over their internal state
- Thus: an agent is a computer system capable of autonomous action in some environment in order to meet its design objectives

An intelligent agent is a computer system capable of flexible autonomous action in some environment

- Typically <u>intelligent agents</u> exhibit 3 types of behaviour:
    — Reactive
    — Pro-active and
    — Social

To continue in the next session.