IST 692: RESPONSIBLE AI

Jasmina Tacheva

Responsibility can never lie with the AI system

- As an artefact, AI **cannot be seen as a responsible actor**.
- We need **chains of responsibility** that can link the system's behavior to the responsible actors.

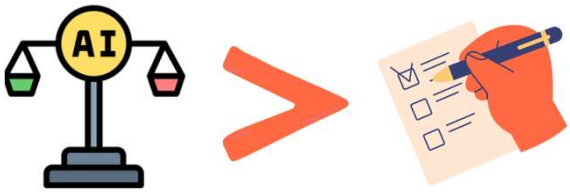
Chains of Responsibility in AI

This course is about…

RESPONSIBILITY:

- being **responsible for the power that AI brings**;
- **our responsibility for the systems we create and use**, and
- about **how**, and whether, **we can embed responsibility** into these systems.

More than "Ethical AI"



Main Challenge

Determine:

- what **responsibility** means,
- who is responsible,
- for what, and
- who decides that?

What is AI?

"[T]he artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving."

'Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' by McCarthy, Minsky, Rochester, and Shannon (1955)

(Artificial) Agents

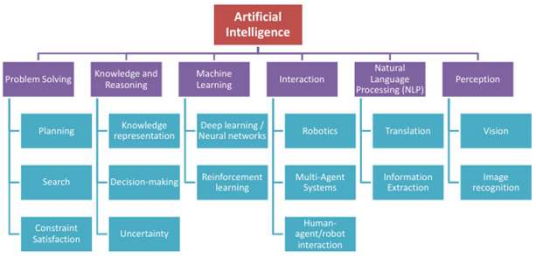
- **Reactivity**: Perceives and responds to changes in the environment
  ~ Adaptability
- **Pro-activeness**: Takes initiative to achieve goals
  ~ Autonomy
- **Sociability**: Engages in communication, cooperation, and negotiation
  ~ Interactivity

What is AI? (Cont.)

Not only about how to represent and use complex and incomplete information but also about:

- how to see (**computer vision**),
- move (**robotics**),
- communicate (**natural language, speech**) and
- learn (**memory, reasoning, classification**)

Main Streams in AI



AI is transdisciplinary

Top-down vs. Bottom-up Approaches to AI

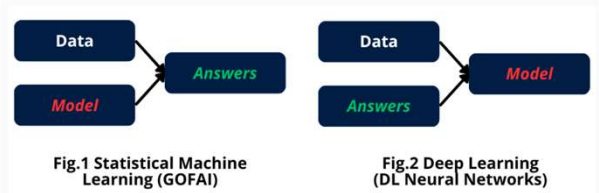
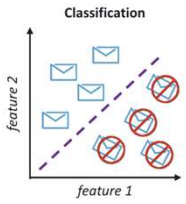


Fig.1 Statistical Machine Learning (GOFAI)

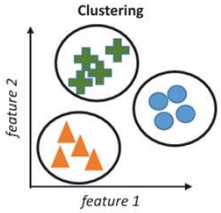
Fig.2 Deep Learning (DL Neural Networks)

Quick Guide to Machine Learning				
Approach		**Unsupervised Learning**	**Supervised Learning**	**Reinforcement Learning**
Objective		Discover structures	Make predictions	Make decisions
Possible Techniques	**Simple domains**	• Clustering	• Regression • Classification	• Markov Decision Processes • Q-Learning
	Complex domains	**Deep Learning** (many-layered neural networks and large datasets)		
Training requirements			Labelled data	Reward function
Example Application		Customer segmentation	Identify spam	Playing a game (e.g. Go)

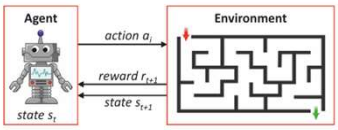
Supervised ML



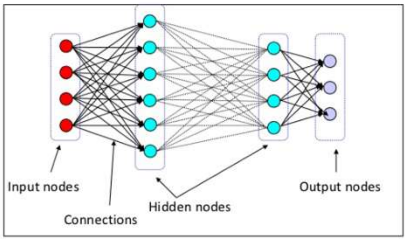
Unsupervised ML



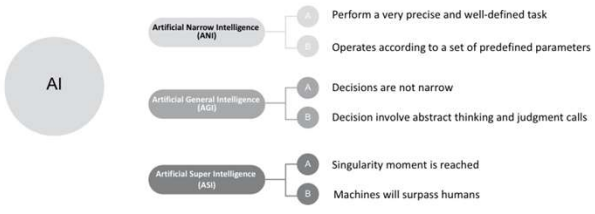
Reinforcement Learning



Deep Learning



Types of AI Based on Scope



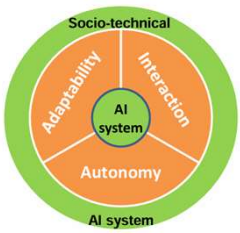
AI Today

The computational capability of interpreting vast amounts of information in order to make a decision.

AI is more than the algorithms it uses!

Muhammad ibn Musa al-Khwarizmi

AI systems are part of a much bigger system



Stakeholders, not just shareholders

Taking Responsibility

- Responsibility **in** Design: Ensuring that development **processes** consider the ethical and societal implications as AI integrates with and sometimes replaces traditional decision-making systems and social structures.
- Responsibility **by** Design: Embedding ethical reasoning capabilities into the **behavior** of artificial autonomous systems.
- Responsibility/Ethics **for** Design(ers): Upholding the research integrity of **developers & researchers** and establishing certification mechanisms to maintain ethical standards.

Key takeaways

- AI has great potential but comes with **significant responsibilities** and ethical considerations.
- As creators of AI, **humans** are accountable for its impacts - both positive and negative.
- AI systems are not just technical artifacts; they are **sociotechnical -** deeply intertwined with societal structures and stakeholders.
- Top-down (**symbolic**) approaches to AI focus on explicit rules, while bottom-up (**connectionist**) approaches rely on pattern recognition from data.
- **Responsible AI goes beyond ethics**; it demands active engagement with societal, legal, and cultural considerations throughout the AI lifecycle.
- ***We*** are responsible for Responsible AI!