

# Artificial Intelligence

## Elective course

Class:	MMT Class 19, 4 <sup>th</sup> semester
Course responsibility/ Instructor:	Thomas Bolander, Associate Professor DTU Compute Richard Petersens Plads, Bygning 322 Phone: +45 4525 3715 E-mail: tobo@dtu.dk

### Course Overview

“Artificial Intelligence (AI) is the science and engineering of making intelligent machines, in particular intelligent computer programs.” This definition of artificial intelligence by John McCarthy, the father of the area, is still valid today, 60 years after it was formulated. Nevertheless, the area is much harder to define and delimit than most other areas of science and engineering, since intelligence is many different things, e.g. social, linguistic or emotional intelligence, and at very different levels.

Correspondingly, AI is many very distinct techniques to mimic different aspects of human intelligence, and at very different levels. This makes it difficult to navigate in the area of AI and understand exactly what it can and cannot do. One of the goals of the course is to clarify the different paradigms, methods and subareas of AI, to provide a clear picture of what AI is, to make it easier to understand the current developments in the area, and to understand the possibilities and limitations of AI methods in general. AI is about mimicking the most essential abilities of humans, our cognitive abilities: perceiving, reasoning, planning, communicating, learning, etc. Hence AI has the potential of transforming our lives, society and businesses in a much more fundamental way than any other technology. Underlying and unifying the different methods in AI is usually the goal of making machines do what hitherto only humans have been able to, e.g. play chess, recognise faces, engage in a dialogue, drive a car or do medical diagnosis. The main source of inspiration for building AI systems that can solve these problems is human problem-solving. In some cases, AI systems are based on a very direct attempt to simulate the fundamental neurological processes of the human brain. In other cases, they are based on more abstract models of human reasoning and problem solving. However, it is important to understand that there still is – and probably always will be – fundamental differences between human and machine intelligence. These fundamental differences are essential to grasp in order to understand where AI techniques can successfully be employed, what kind of human tasks can be replaced or enhanced by the use of AI, and how AI is going to affect the future job market. These aspects will all be thoroughly covered in the course.

The World Economic Forum Global Risks report of 2017 deem “artificial intelligence and robotics” to have both the highest potential benefits and the highest potential negative consequences among all emerging technological areas. The course will address both the benefits and the potential negative consequences, including discussions of the legal, ethical and philosophical aspects of AI.

The course will be held by DTU faculty who are active researchers in the area of AI. The format is lectures, group discussions and group exercises. The course is concluded by a written take-home assignment which constitutes the assessment in the course.

## Learning Objectives

- Explain concisely the scope of AI, its potential for society as well as its limitations
- Discuss contemporary applications of AI from technical, legal, ethical and economical perspectives
- Describe in overall terms the main paradigms of AI and explain their respective strengths and weaknesses
- Describe in overall terms the most prevalent methods in current AI and their respective application areas, strengths and weaknesses
- Explain the meaning of, and difference between, concepts often used to describe AI or parts of AI, including deep learning, machine learning, big data, cognitive computing, and artificial general intelligence.
- Explain the current and expected future impact of AI on different domains like transport, healthcare, education, entertainment, etc.
- Given a task currently solved by humans in the student's own area of work, analyse which methods of AI could potentially solve this task, if any, and to which extend
- Given a task currently solved by humans in the student's own area of work, analyse the advantages and disadvantages of using AI systems instead, in terms of quality of the task solution, efficiency, ethical, economical and societal consequences
- Critically analyse and discuss news media coverages of AI
- Describe the main trade-offs in designing AI systems in terms of generality vs. scalability, flexibility/learnability vs. predictability/explainability, etc.
- Clarify the impact of increase in available data and computing power on AI
- Explain the main characteristic differences between human and machine intelligence, and the potential in human-machine collaboration

## Assignments

The course has two assignments, "Critical review" and "AI in my area". You need to prepare for both before the course starts, see the short descriptions below.

### Critical review

This is a group assignment in groups of 3 to 4 students (to be formed at the beginning of the course). In this assignment, you will critically review a mainstream news media article on AI. The group will present a 10 minute slide show in the afternoon session on Saturday, followed by 10 minutes Q&A from the other course participants. Each student should select a relevant news media article before the start of the course (see notes on preparation below), and the group will then during the course select one of these articles and prepare a slide show critically analysing it.

*Preparation before course start:* Each student should individually find a fairly recent (within the last year) mainstream media article on some aspects of AI, e.g. on technological developments, on opportunities or threats of AI, on philosophical, ethical, legal, economical or societal aspects of AI. You can search for such articles through the web pages of your favourite news media or for instance use the NewsFinder site for AI (<http://aitopics.org/news>) or the IBM Watson New Explorer (<http://news-explorer.mybluemix.net>). Prepare a 2 minute presentation of the article to your group members (informal presentation, no slides).

## AI in my area

This is an individual written assignment to be handed in at most 14 days after the last day of the course. In this assignment, you should analyse a task or set of tasks in your own area of work that you expect could be fully or partly automatised using AI methods. You should briefly analyse which of the AI methods presented during the course might be relevant to use, to which degree the tasks can be automatised, analyse aspects of human-machine collaboration, and analyse the advantages and disadvantages of automating the tasks (including ethical, economical and societal consequences). You are free to choose anything from a very narrow, repetitive and well-defined task that AI can easily handle (or is expected to) to very ambitious AI systems that could replace your entire job (or someone else's).

*Preparation before course start:* Select a relevant task or set of tasks in your area of work (ideally from your own workplace) that you would like to analyse further with respect to full or partial automatisation using AI. Prepare a 1 minute presentation describing the tasks to your group members (informal presentation, no slides).

## Sessions

### Introduction to AI & Sub-symbolic AI

Time: Friday, February 2, 08:30 – 12:00

Topics: **Introduction to AI:** The concept of AI, characteristics of AI today, simple and hard problems in AI, human intelligence vs machine intelligence, symbolic vs subsymbolic AI.  
**Sub-symbolic AI:** Artificial neural networks, deep learning, genetic algorithms.

Readings: 1) Chapter 1 (Introduction) of Russell & Norvig “Artificial Intelligence – A Modern Approach”, 3<sup>rd</sup> edition, Pearson 2016.  
2) Pages 1–17 + 50–52 of “Artificial Intelligence and Life in 2030 – One Hundred Year Study on Artificial Intelligence”, Stanford University, September 2016.

### Symbolic AI & Current trends in AI

Time: Friday, February 2, 13:00 – 16.30

Topics: **Symbolic AI:** Search and planning, knowledge representation, reinforcement learning.  
**Current trends in AI:** Artificial general intelligence, combinations of symbolic and sub-symbolic approaches, explainable AI.

Readings: No additional reading.

### **Hard problems in AI & The impact of AI**

Time: Saturday, February 3, 08:30 – 12.00

Topics: **Hard problems in AI:** Linguistic and social intelligence.  
**The impact of AI:** Impact on different domains like transport, service robots, healthcare, education, etc.; impact on job market and society.

Readings: Pages 18–49 of “Artificial Intelligence and Life in 2030 – One Hundred Year Study on Artificial Intelligence”, Stanford University, September 2016.

### **Ethical, legal and philosophical aspects of AI & AI in the media**

Time: Saturday, February 3, 13:00 – 16.30

Topics: **Ethical, legal and philosophical aspects of AI:** Robustness, machine bias, strong vs weak AI, human-in-the-loop, autonomous weapons, threats from AI to society, legal challenges, ultrainelligence.  
**AI in the media:** Group presentations of critical reviews.

Readings: Chapter 26 (Philosophical Foundations) of Russell & Norvig “Artificial Intelligence – A Modern Approach”, 3<sup>rd</sup> edition, Pearson 2016.