IN5620 Interaction with AI and autonomous systems

Lecture notes 22 August 2024

Updated 22 August 2024

1. Introduction 1
2. Interaction with AI today 4
3. Autonomy 6
4. Intelligence and consciousness? 7
5. Humans and machines 8
6. Interaction with AI - a pragmatic discipline - a pragmatic view 9

1. Introduction

Welcome.

We are two lecturers in this course; Asbjørn from Sintef and Jo from IfI. In addition, Mike is the teaching assistant.

We have divided the course into three different modules. The three modules are:

- What we talk about when we talk about interaction with AI and autonomous systems.
- Design of interaction with AI.
- Reflection and writing/discussing/reviewing/thinking.

We start with an introduction and background module, and a guiding question: *"What do we talk about when we talk about interaction with AI?".* Module two is about how to design interaction and AI interfaces. The third module is focusing on writing and working with the projects and individual reports.

There is a curriculum for the two first modules. This consists of articles that we will use throughout the course. The curriculum will be updated as the course proceeds.

We believe that nobody is an "expert" in the field of "interacting with AI and autonomous systems" today; it is an open and indeed evolving field. Our colleague here at UiO, Anders Kluge likes to say that the computer – and the use of the computer is "i støpeskjeen"; a

Norwegian idiom indicating that the computer, and the use of the computer, is evolving. Computing and the use of computers 60 years ago are both very different and very similar to computing and the use of computers today. It is similar from the perspective of a basic interaction model:

a machine gets input from human(s) in a context - then
the machine process – or do something with the input - then
the machine provides some output to the human (s).

Another similarity is about representations. The input to the computer, the output from the computer and the programs that the computer runs are represented by numbers, it is symbols.

It is also different: today the machine gets input from a variety of sensors and large data sets, the algorithms and machinery that does the processing have evolved – and the input and output mechanisms have changed (images, sounds, “patterns” as input and generated patterns as output).

Even though nobody is an expert in this field, we all have different experiences, thoughts and interests in the field of “interacting with AI and autonomous systems” in this class. We believe that You all have unique interests and experiences with various areas that are valuable for all of us. We hope that you will engage with the course, discuss, investigate, present in class, think, reflect, do and write a lot - and by this learn something valuable about “interaction with AI and autonomous systems”.

There will be both individual work and group work in this course. At the end of module 1, you will hand in the first group assignment; a wonder document. At the end of module 2, the mid-term report is due, and the final report from the groups is due at the end of module 3. This means that at the end of the course, you will have handed in the group assignment in three iterations.

The individual assignment is structured as follows. For the two first modules, you will get tasks to discuss and do, and write about. You will at the end of the course have handed in the individual assignment in three iterations.

For both the group assignment and the individual assignment, we will have a peer review process; i.e., that you read and comment on each other's assignments. After each delivery, you will read and comment on another delivery, in a round-robin fashion based on a list. For the group delivery, another group is reviewing, and for the individual deliveries, another student is providing feedback.

So, one assignment in the course is to provide review written reports. Equally important is to make use of the reviews that you receive - and learn and do something about the comments that you get. For delivery 2 and 3 of both the individual assignment and the group assignment, it is expected that you write in the appendix what you have done with the feedback you have received.

Note that the peer review structure is a condition for attending this course. To do the course, you will be given the opportunity to learn from other participants' writings, and the project work done in the groups.

We will use the Vortex system during the course; for messages, for the overview of the plans and deadlines - so all communication from "the course" to you will be found there. This is an open platform that you hopefully are familiar with. It is open in the sense that what is published there is available to all other students. For the deliveries, each group will get a folder - and you can post and deliver your assignments there - both the assignments themselves - and the feedback. You are welcome to use other platforms for discussion, collaboration and communication for communication within groups.

All groups will also present the current status of their project and one article during the course in class. This is done so that we can learn from each other, and discuss the projects – and the articles in class.

A requirement for the course is that you have done all the assignments. In the end, there is an oral presentation.

Finally, some words about workload for the course:
This course is 10 ECTA (European Credit Transfer and Accumulation System)
Lectures/session, 2 hrs every week, including a break.
Group work, x hrs every week.
Individual work, y hrs every week.
2+x+y=10-15 hours every week

If we say that 30 ECT is 30-hour work week, then 10 ECT is 10 hours work pr. week. If we say 30 ECT is 45 hours every week, then 10 ECT is 15 hours every week.

The next section is going into Module 1 of the course.

2. Interaction with AI today

The words Interaction, Artificial Intelligence and Autonomous are key concepts that are used both to describe “humans, like you and me” and “things, like a computer or a service”. Traditionally interaction, intelligence and autonomy are concepts to describe some characteristics of the human condition. Today, we use the very same concepts, and the very same words to describe computers and computing systems.

What follows is a transcription of a session with an Android phone, where: J=Jo, G=Google Assistant.

J: Please set an alarm for 45 minutes.
G: Done. Your alarm is set for 7:25.
J: Thanks
G: That's what I am here for/I was literally made for this kind of thing (and a smiley on the displayed on the screen)/You’re welcome.
J: Please delete the alarm.
G: You can do that in the app

You are all familiar with setting an alarm; most of us do it frequently. Traditionally, we have used buttons and dials on physical wristwatches or alarm clocks, text commands in Linux, icons and lists in a calendar program or app - or as you saw her with your voice input via a microphone and voice output from a loudspeaker in concert with screens.

These are different *dialogue systems*, where interaction takes place through or via an interface of some sort. When using voice, we may have a dialogue system based on *conversational interaction*; where we are talking and listening to the computer as if it was a conversation with a fellow human being. We will in this course both in module 1 and in module 2 get back to conversational interaction, especially in module 2 where we will work with chatbots. The word chatbot is a composed word of chat and robot. We will also get back to what we talk about when we talk about robots more in module 1. What do we mean – and how do we talk about Intelligence, autonomy, robots, chatbots, communication, interaction and conversations today? What do we talk about when we use these words today – when they are used both from a “machine” perspective and a “human” perspective?

Then, back to setting the alarm with a conversational interface, and with voice commands. It is fairly recent that machines can be used in this way in everyday life – with spoken commands or sentences. Speech recognition is one instance of pattern recognition that during the last 10 years has improved in performance and accuracy – and today speech recognition is working well in many everyday situations. WIMP (windows, icons, menus,

pointer) and the QWERTY keyboards have been important input devices – and now the microphone and corresponding pattern recognition system are complementing this.

Another situation of use that many engage in when interacting with a clock, is in the evening when going to bed. Then to set an alarm for let's say 8 hours, and get feedback from the machine, visually or orally, that this will be at 06:28 - gives both an assurance that the alarm is set ok - and we can judge if it is too late - or an ok time to be woken up. For many, this feedback from clock machinery is sometimes valuable, but is it an “intelligent” interface because it works like this? Is it a “smart” system? Is the clock machinery autonomous?

Following up on the example a bit further, when trying to delete the alarm, I did not succeed in the dialogue, conversation – or interaction with the machine. I was sent to an “app”, where I was asked to find the alarm, find the appropriate places on the screen and press this. It was not deleted directly after the command or part of the conversation “please delete the alarm”.

This, relatively concrete, mundane and unremarkable example of a use situation can be investigated further - analysed, talked about - and it is possible to work with understanding this interaction to design new ways of doing this. If the “media or alarm audio level" is set to 0, or the machine is on “silent”, it would be good to get feedback on this in some situations for example. Or if we by applying insights into Universal Design can make this kind of functionality accessible to more users, in more situations, for more human activities?

When selecting a theme - and questions to work with in the group work; we encourage you to be specific - and find phenomena that it is possible to work with hands-on. Today, AI and autonomous systems are “infused” or built into many applications and services, so there are possibilities for finding something interesting. Perhaps the example of how interacting with a timepiece, a mundane clock is done - and can be done - can inspire you to look for other interaction settings.?

Some of the areas to look for inspiration for projects are:
Spoken language, or images as input/output - detection and use by the machine. For example - when writing a text, is it possible to do this today by speaking – or having a conversation with a machine instead of writing on a keyboard/screen? In what way do the work process and work practice of writing change? When reading and annotating articles, what practices exist in switching between reading text and listening to the text?
Other patterns as input for the machine; movement patterns etc.
Patterns as output from the machine-generative AI in human-readable form; image, text or sounds? In what way can generative systems simulate drawings or texts, plans or whatever in ways that are meaningful and helpful for you?

Integration/infusion of AI system into -- libraries for reference finding, text editing, music recommendation - and many other examples.
Inclusion and access in new ways for specific users….

3. Autonomy

There is another issue that can be illustrated with the clock - apart from the situated and specific interaction taking place. A clock, or a timepiece, or a time system is an example of **autonomous** technology. It does something "on its own". The primary function is handled, or operated **autonomously** by the clock machinery, without human intervention. When the machinery is working, it just goes on doing what it does without any input or control from humans, or users. Even underwater, in the sky - or here among us.

This clock technology can be understood and explained by dividing it into some modules that work together so that it is a clock. The modules are:

- An oscillator; something with a regular frequency.
- A counting mechanism; something that counts the oscillations.
- A display/interaction module that enables presenting the time and setting the time.

In the seminal book Computer Power and Human Reason, from Judgement to Calculation, Joseph Weizenbaum uses the clock as an introduction and explanation of autonomous technology. The first chapter of the book is called "on tools", where he describes the evolution of important tools that we humans have made - both the tools that make us do different *physical activities* like cutting down trees and chopping wood with axes, shooting pray over distance with guns, and *sensory instruments* like the microscope and telescope to let us discover more things in the world we live in. Among these tools, the clock is presented as having had a specific role, as the first **autonomous technology**. The clock has a "model" of the world built into it (from sunrise to the next sunrise – it is 24 hours, each hour is divided into 60 minutes, which again etc…), and it operates on this model.

It is perhaps hard to see this today, since we are familiar with the clock in our daily lives - but it is possible to imagine life before the clock - and see clock machinery as an autonomous machine – or as an Artificial Intelligence, and AI? Imagine living 1000 years ago, before any timepieces were around. It is possible to imagine that you had a fine sense of what time of day it was based on your reading of the sun. With the clock, your relationship with the sun and the position of the sun did change - as well as opportunities for coordinating with fellow humans in new ways.

This is an example of a wall clock. It can be used to tell the time. This particular wall clock is set **automatically** from a very precise master clock in Germany, over the air. Much like the clock integrated into the smartphones we use today. You do not have to do anything to make it work, it works autonomously. And it brings its logic with it. It does something very much more precise than any human can do, indicating "clock" time accurately. And it does it "on its own", even "adjusting and correcting itself" from a master clock machine somewhere in Germany.

At 12:15 we started today. We are all users of this autonomous technology, and it enabled us to meet "here". When it indicates 14:00, we will end the lecture. It structures the social setting here, and we are of course all very familiar with this way of meeting - it is a mundane, familiar, everyday technology that is one of the conditions for the possibility of living in the world we live in today. The point I want to make here is that it is an **autonomous technology** and that autonomy is at the very core of many technology developments. Autonomous weapon systems, autonomous cars and vehicles, and autonomous conversational agents are examples of areas where AI is at work today.

We are humans, and autonomy is used to describe one of the conditions of being human. Integrity and freedom are other words, often used to describe some of the same phenomenon. What does happen when we mix these two types of autonomy? What happens when machine autonomy and human-autonomy meet?

4. Intelligence and consciousness?

Intelligence and consciousness are different. In general terms, intelligence is the word we use for talking about the ability to solve problems, find alternative ways to problems and attain goals. Since the concept first appeared, it has been studied, discussed and used in many different areas - also within the field of Human-Computer Interaction (HCI).

Intelligence is a double-edged sword. Intelligence enables us to invent and make many new tools, such as writing equipment. These tools again do affect and shape us into who we become and what we can do in this world. Today, computers simulate intelligence, i.e. problem-solving in different ways.

We know that we are conscious because we feel and experience the world. We cannot scientifically "prove", or "measure" consciousness. But of course, we have social conventions that say consciousness is something real in humans - and not in for example a hammer or a computer.

We can program computers to imitate humans and simulate humans (movement, voice, etc) to pretend it is a human with intelligence and consciousness. The important words here are "pretend" and "simulate" - it is artificial. This is seductive for humans, for we want to be listened to, and we want to build relationships - and this may be simulated in machines.

The machine, as far as we know, does not feel or experience anything - and hence does not have any consciousness. However, added to the discussion about AI today are possible future computers with Artificial General Intelligence (AGI), Artificial Narrow Intelligence (ANI) and Artificial Super Intelligence (ASI).

5. Humans and machines

We are humans, not machines. We humans are feelings, and we sense and experience in more or less meaningful ways. You may think that “today I am like a machine”, or being on “auto-pilot”, and even as a “slave to the machine”; but we “are not” machines even then. Maybe the sentence “I feel, therefore I know that I am” is a more precise statement than “I think, therefore I am?”. When for example experiencing suffering in yourself - or others, this is evident. Today, when we talk about machines that “think” and that are “smarter than us, more intelligent, and even win chess games”, it is a small step further to think and talk about machines as they were “humans”.

Of course, there are many similarities between humans and machines. Analytically, we can select some areas where there are similarities between humans and machines.

“Physical”; human bodies and "hardware" for computers.
"Sensors", vision with eyes for humans and cameras or sensors for the computer.
"Effectors", the voice for humans and loudspeakers for computers.
"Reasoning", thinking for humans and calculations for computers.
“Emotions”, feelings for humans and simulations/representations of emotions for computers.

Both the established research fields *evolutionary computing,* and *bio-inspired computing* deal with learning from humans and other species that have evolved, making models of what is done by humans and other species - and then engineering and programming it.

There are some differences between machines and humans. Weizenbaum, Dreyfus, Suchman, Winograd, Agre, Searl, Harari and others have gone to great lengths to discuss important differences. One topic that often is described is about “making sense”, or “finding meaning”, or “feeling” as a basic difference - where we humans seek to make sense and experience the world we are in - whereas machines do not. We experience being-in-the-world, being present here, now - whereas machines do not. It makes sense for me to be here now, for example, to do my job as employed by the university. For computers, it does not, as far as I know. You have knowledge about calculations and nutrition for example, whereas machines do not possess this kind of “knowledge”. You learn new things every day. Machines do not learn, at least not in the way we have historically thought about learning. This lecture may make sense (or not) for you. This lecture does not make sense or mean

anything for the computer. You may understand the equation 1+1=2. The computer does not understand 1+1=2. We can experience suffering; the machine can not suffer.

And here, we are at an important issue with ways of talking and thinking about machines and of us humans. We use many words, concepts and categories that historically meant something about being human and apply it to systems and machines. Knowledge, understanding, communication, interaction, perception, intelligence, learning, sensing, agency, intention, dexterity, curiosity, affections, autonomy - these are all words that we have used to talk about human conditions, being a human.

Within AI, many of these concepts and words are used merely as metaphors, but in other instances, they are applied more literally. What is important for this module in the course is to discuss what we mean when using these words and be open to discussing it. The concept of *anthropomorphism* and *zoomorphism* are ways of talking about this.

A pragmatic view is that we apply and use metaphors in computing - and when talking about using computers. The "desktop" on "Windows" is not a desktop, and a “web page” is not a page in the traditional, original sense. Navigating the web is different from navigation and wayfinding in the mountains or at sea. And machine learning is different to human learning. In what ways are humans and machines different? What do we mean when we say that machines learn? And, what is learning for us humans in the age of machine learning?

6. Interaction with AI - a pragmatic discipline - a pragmatic view

AI is an established research topic in computer science and industry. “Interaction with AI and autonomous systems” is not an established research field; neither in computer science nor in industry. In the following, we will say something about the AI field as a pragmatic field.

In 2016 during the annual Google I/O conference, at the keynote opening session by the chief executive officer Sundar Pichai, it was announced that the old slogan “mobile first” is replaced by “AI first”. This was repeated and elaborated upon in later I/O conferences. Google is communicating to us, developers and consumers, that AI will affect the products and services they work with in many different ways and many different areas, such as in the health domain. This is an important message since it is from one of the largest ICT enterprises of our time and since services and products from this company are used by so many different people, in so many different situations, in so many different activities. Other examples are Siri from Apple, IBM Watson, Alexa from Amazon and Cortana from Microsoft

and recently chatGPT from OpenAI and Microsoft, Perplexity….. These are all said to be "intelligent" in some way, and powered by AI.

Microsoft, on their official website, writes a slogan for AI: “Amplifying human ingenuity; we’ re excited about the opportunities that AI bring to organisations and people.”

One of the many companies making platforms for AI, viv.ai has this slogan:
“Radically simplifying the world by providing an intelligent interface to everything.” - and “intelligence as a utility”. The way the company present itself on the web is, on the “front page” (it is very hard not to use the metaphor “page” here, of course, it is not a “page” literally (unless you have printed it onto a page), but text and images on a display in front of you):

“Viv is an artificial intelligence platform that enables developers to distribute their products through an intelligent, conversational interface. It’s the simplest way for the world to interact with devices, services and things everywhere. Viv is taught by the world, knows more than it is taught, and learns every day.”

What does this mean? It sounds incredible, that something is distributed through an intelligent, conversational interface, and that “the platform..knows more than it is taught, and (that) it learns every day”. This can be interpreted in a pragmatic way that the company is providing a platform that applies machine learning, conversational interfaces and knowledge systems or big-data systems. It is not “really” about learning, knowledge or conversation - as we have used these words and concepts traditionally.

Mass media is discussing AI. It is possible to find statistics on the frequency of articles, and documentaries on this topic over the years since the term “AI” was coined. Indeed, it could be interesting to see the number of mass media articles over the years, we know that during the AI winters with little funding, AI got minimal attention from mass media. Today, there is much discussion of AI in mass media. Many people are engaged with different emotions, such as great hopes and great fears, and news about the hopes we have for what AI can do for us - and fears we have for unintended or unseen consequences are there.

And lastly, AI has been, is - and probably will continue to be explored in films, books and games. To name a few local and global: the film 2001 A Space Odyssey from 1968 directed by Stanley Kubrick and inspired by a short story by Arthur C. Clarke. The Norwegian TV series Blindpassasjeren by Jon Bing and Tor Åge Bringsværd from 1976 is another example.

The stories about what AI is/is not are worth working diligently with – from narratives in the news, films and games – and from the ICT industry.

Stories and narratives about AI are high on the agenda among technology companies, in scientific research, in media worlds and portrayed in films, books and games. On a macro level, on the political and social agenda, on a meso level, organisations for example and finally on a micro level - you interacting with a system.

What do we talk about when we talk about Human interaction with Artificial Intelligence and autonomous systems? This is the guiding question for the first module.

Next week: we continue with this question and will go into AI and Universal Design. We will also continue with forming groups.

See you next week