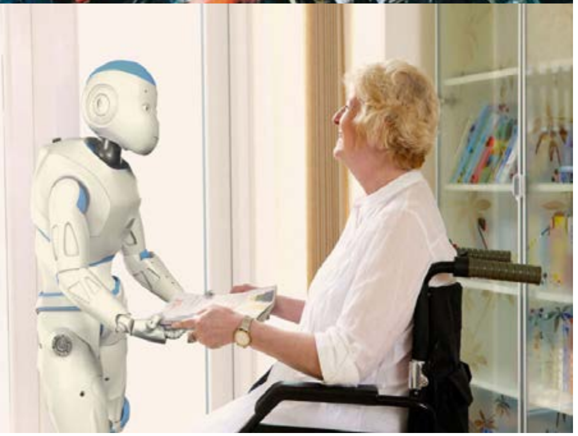
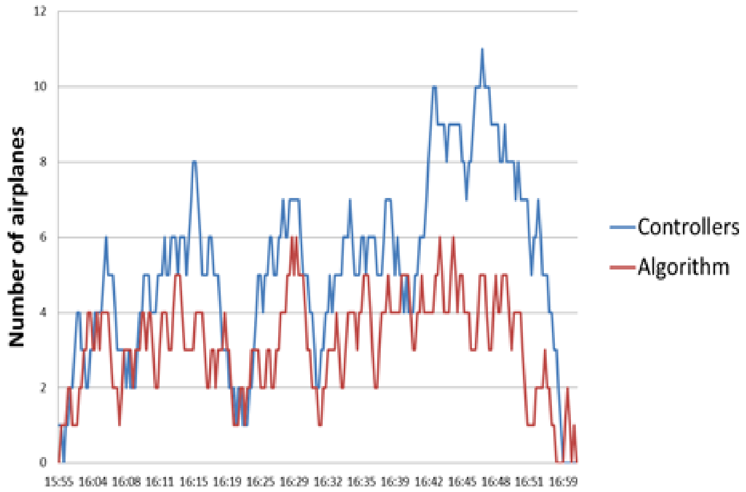



Interacting with AI
Module 3
Working and living with AI

Amela Karahasanović
amela@sintef.no

Senior Research Scientist, SINTEF Digital
Associate Professor, IFI, DIGENT

Decision support in ATM
Evaluation methods in HCI
User Experience
User behaviour



1 / 9

Feature Article:

DOI. No. 10.1109/MAES.2019.2916039

Can Holistic Optimization Improve Airport Air Traffic Management Performance?

Amela Karahasanović, Aslak W. Eide, Patrick Schittekat, Hans Erik Swendgaard, Krystsina Bakhrankova, Dag Kjenstad, Carlo Mannino, SINTEF, Oslo, Norway
Theodor Zeh, Volker Grantz, FREQUENTIS AG, Vienna, Austria
Carl-Herbert Rokitansky, University of Salzburg, FB Computer Sciences Institute, Salzburg, Austria
Thomas Gräupl, Institute of Communications and Navigation, Wessling 82234, Germany

INTRODUCTION

Air transportation is an important factor in the economic growth of the European Union; however, the current system is already approaching its capacity and cost limits, and therefore needs to be reformed to meet the demands of further sustainable development [1]. According to the European Commission, airspace congestion and the delays caused by it cost airlines between €1.3 and €1.9 billion a year [2]. Several research initiatives have been launched to address air traffic management (ATM) challenges. The Single European Sky ATM Research (SESAR) program—a joint effort of the European Commission, EUROCONTROL, air navigation service providers, and the manufacturing industry—aims to define, develop, and deploy what is needed to increase the ATM performance and build Europe's intelligent air transport system.

Similarly, in the United States, the Next Generation Air Transportation System (NextGen) is the Federal Aviation Administration-led modernization of United States' air transportation system to make flying even safer, more efficient, and more predictable.

Reducing gridlock, both in the sky and at airports, is one way to improve the efficiency of the air transport system. However, according to Anderson and Milutinovic, [3] recent improvements to enroute capabilities have caused a shift in air transport systems, meaning bottlenecks at the airport are now the primary concern.

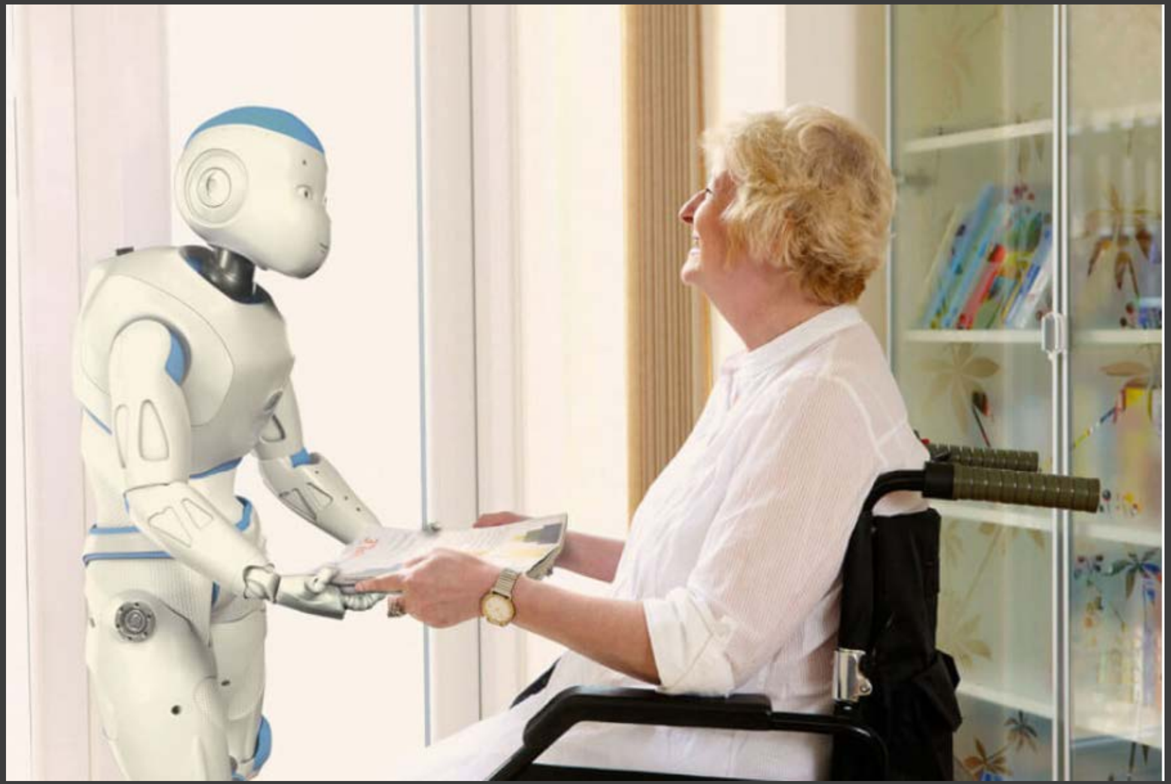
As such, research on mathematical optimization methods to support decisions near and at the airport is of great interest. Marin and Salmerón [4], [5] were the first to demonstrate a taxi planning optimization tool, which minimized the overall taxi time at the Madrid-Barajas airport based on a space-time multicommodity network with capacity constraints. Stiverson and Rathinam [6] addressed the runway-queue management problem of the Dallas/Fort Worth using fast search heuristics based on k-exchange neighborhoods. Erzberger et al. [7] proposed an arrival-sequencing algorithm integrated with separation management and weather avoidance within the wider

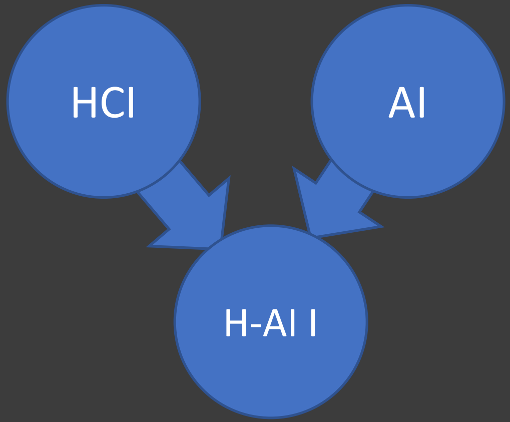
Why am I interested in this?

Objectives of Module 3

Understanding of challenges related to use of AI based systems in everyday life and at work

- How to evaluate them?
- When and how to use them?
- How to integrate them in our life?





Module 3 overview

Evaluation of interaction with AI (3.1&3.2)	**5th of September**
Human - AI partnership (3.3&3.4)	31st of October
Lessons learned from studies of human – AI interaction (3.5&3.6)	7th of November
Writing workshop	14th of November

Module 3 - assignments

Individual assignment

- Building on and extending the individual assignment in Module 2. Start after the end of Module 2 – finish 22nd of November

Group Assignments and tasks

- Building on and extending the group assignment in Module 2. Start after the end of Module 2 completed – finish 22nd of November
- Task 3#1 (lesson 3.1&3.2) – task on evaluating interaction with AI. Start after 3.3&3.4 session, finish in one week, include as Appendix 3 in the final report
- Task 3#2 (lesson 3.3&3.4) – task on human-machine partnership. Start 31st after 3.5&3.6, finish in one week, include as Appendix 4 in the final report

Plan for today

- Evaluation – why, what and how to evaluate
- Shifting the focus of AI evaluation - User Experience, trust and values
- Several small tasks for group work

Why to evaluate?

Airport passport controll

Your turn - Task 1

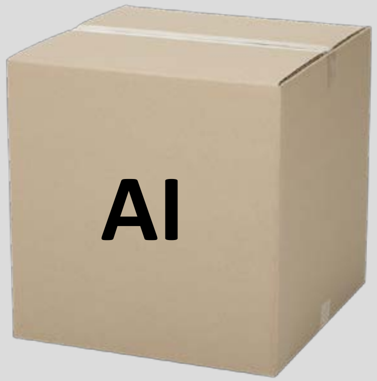
Find a video which illustrates well some of the problems that might appear when we interact with AI, an "AI/robot goes crazy" example.

Group work – 10 minutes

- What was the problem?
- Could it be solved differently?
- Could the problem be discovered earlier?
- What are the possible consequences?

What to evaluate?

What to evaluate?



From the previous lecture

- Narrow intelligence
  - AI that is good at performing a single task
- AI, Machine Learning, Deep Learning

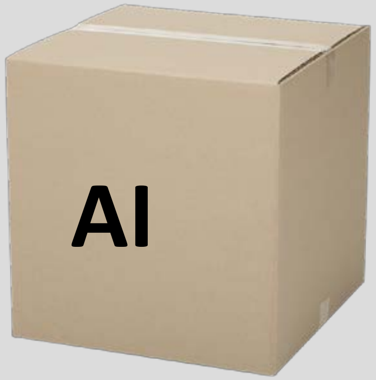
What to evaluate?

Task

Application domain

Context of use

Users



Definitions of AI

- McCarthy (2007) – "AI is the science and engineering of making intelligent machines" –> intelligence test
- Minsky's (1968) - "AI is the science of making machines capable of performing tasks that would require intelligence if done by humans"

  -> task-oriented evaluation
- **AI effect** (McCorduck 2004) - tasks are not considered AI problems any more once they are solved without full-fledged intelligence

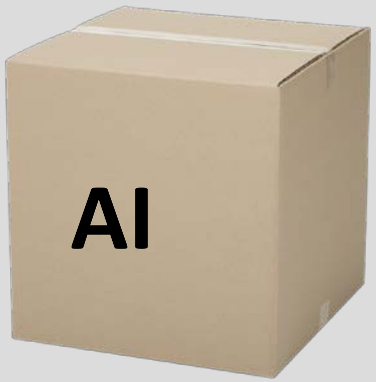
What to evaluate?

Task

Application domain

Context of use

Users



What to evaluate?

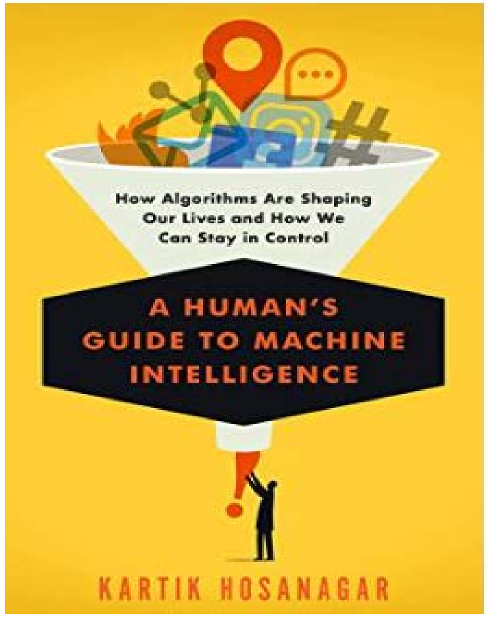
- *AI systems* - robots, chatbots, social robots, AI agents, *self-driving car*
- *AI components* - techniques, algorithms, methods or tools, *camera of the self-driving car*

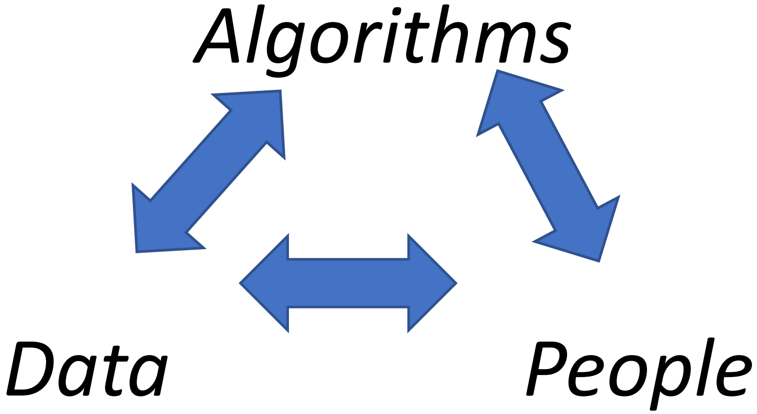
- *Systems evaluates as they are – components according to a specification and how they the serve the system*
    - *Formula 1 engine not appropriate for a family car*

(Hernández-Orallo, 2017)

What are we actually evaluating?

"The results of the algorithmic systems can be attributed to their underlying data, their mathematical logic, and the ways in which people interact with their decisions and suggestions" (Hosanagar, 2019)





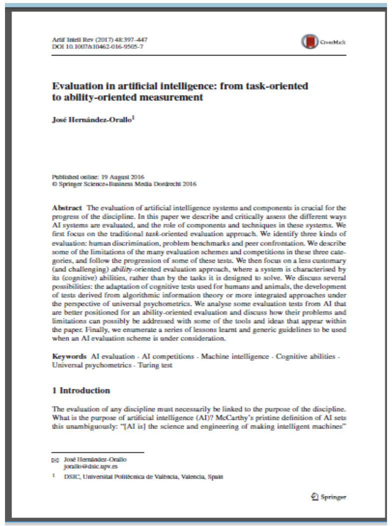
Consider a recommendation engine!

How to evaluate?

From task-oriented evaluation to ability based evaluation

Hernández-Orallo (2017)

- *Task*-oriented evaluation approach
  - Specialized AI systems
- *Ability*-oriented evaluation approach
  - General-purpose AI



Artif Intell Rev (2017) 48:397–447
DOI 10.1007/s10462-016-9505-7

Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement

José Hernández-Orallo[1]

Published online: 19 August 2016

Abstract The evaluation of artificial intelligence systems and components is crucial for the progress of the discipline. In this paper we describe and critically assess the different ways AI systems are evaluated, and the role of components and techniques in these systems. We first focus on the traditional *task*-oriented evaluation approach. We identify three kinds of evaluation: human discrimination, problem benchmarks and peer confrontation. We describe some of the limitations of the many evaluation schemes and competitions in these three categories, and follow the progression of some of these tests. We then focus on a less customary (and challenging) *ability*-oriented evaluation approach, where a system is characterised by its (cognitive) abilities, rather than by the tasks it is designed to solve. We discuss several possibilities: the adaptation of cognitive tests used for humans and animals, the development of tests derived from algorithmic information theory or more integrated approaches under the perspective of universal psychometrics. We analyse some evaluation tests from AI that are better positioned for an ability-oriented evaluation and discuss how their problems and limitations can possibly be addressed with some of the tools and ideas that appear within the paper. Finally, we enumerate a series of lessons learnt and generic guidelines to be used when an AI evaluation scheme is under consideration.

Keywords AI evaluation · AI competitions · Machine intelligence · Cognitive abilities · Universal psychometrics · Turing test

1 Introduction

The evaluation of any discipline must necessarily be linked to the purpose of the discipline. What is the purpose of artificial intelligence (AI)? McCarthy's pristine definition of AI sets this unambiguously: "[AI is] the science and engineering of making intelligent machines"

José Hernández-Orallo
jorallo@dsic.upv.es

1 DSIC, Universitat Politècnica de València, Valencia, Spain

Springer

From task-oriented evaluation to ability based evaluation

Hernández-Orallo (2017)

- *Task*-oriented evaluation approach
  - Specialized AI systems; clear goals: speech recognition, game playing
  - Does the system perform the task
- *Ability*-oriented evaluation approach
  - General-purpose AI: artificial pets, assistant, smartbots…variety of tasks
  - Abilities: verbal abilities, learning abilities, motion abilities

How to evaluate?

- AI applications: computer vision, speech recognition, music analysis, machine translation, text summarisation, information retrieval, robotic navigation and interaction, automated vehicles, game playing, prediction, estimation, planning, automated deduction, expert systems
- Task-oriented
  - human discrimination – assessment by or against a human – it is being subjective
  - problem benchmarking – assessment against a set of problems - difficult to define good sets, for example self-driving car perform well in California, but not in Nord Norway
  - Peer- confrontation – competition against another system; the results relative to the oponents

How to evaluate?

- Ability-oriented evaluation
  - Psychometrics (IQ tests and similar) – not really adequate, "ability" very anthropocentric not reflect the diversity of AI systems

- Generic guidelines
  - Specify the set of systems to be evaluated, the set of possible tasks, describe the similarities between the tasks

Can we use what we already know about interaction evaluation?

Usability engineering

Activities aiming to improve the ease of use of an interface

- Expert-based testing *(usability inspection)*
- Automated testing *(usability inspection)*
- User-based testing *(usability testing)*

Expert-based testing

Interaction Design course

- Structured inspections done by interface experts
- Before tests with users
- Confusing wording, inconsistent layout, obvious flaws
- **Heuristic review**
  - **Compare interface with the rules**
- Consistency inspections
  - Series of screens or web pages inspected
- Cognitive walkthrough
  - Experts perform the tasks (high-frequency and important/seldom)
- Guidelines review
  - Web Content Accessibility Guidelines

Eight Golden Rules of Interface Design

July 15, 2008 by web2usability

Eight Golden Rules of Interface Design

As a result of Interface Design Studies, Ben Shneiderman proposed a collection of principles that are derived heuristically from experience and applicable in most interactive systems. These principles are common for user interface design, and as such also for web design.

1. Strive for consistency.
2. Enable frequent users to use shortcuts.
3. Offer informative feedback.
4. Design dialog to yield closure.
5. Offer simple error handling.
6. Permit easy reversal of actions.
7. Provide the sense of control. Support internal locus of control.
8. Reduce short-term memory load.

Your turn - Task 2

Use Microsoft guidelines for human AI interaction to quickly evaluate Siri or Google assistant

Group work – 10 minutes

Groups 1 and 2-> M1

Group 3 -> M9

Group 4 and 5 -> M11

https://aidemos.microsoft.com/guidelines-for-human-ai-interaction/demo

Shifting the focus of AI evaluation

Can AI play chess better than humans?

-> How can AI empower humans?

Future AI systems should focus enhancing human cognitive capabilities and channelling human creativity…incorporating trust, ethics, and human values

Global effects of a 'local' optimal solution
Values, ethics, privacy and security as a core design considerations
Embedding ethics and values into AI system

(Lukowicz, Slusallek, 2018)

Example: User Experience with robots

- Context: factory
- Two types of robots, one within a safety fence
- UX questionnaire (23 respondents)
- Study – a year and the half
- Compared two types of robots
- Used UX factors in HRI for the evaluation
- UX over time
- Covered aspects: cooperation, perceived safety, perceived stress, perceived usability, general UX
- Conclusion – things take time; ratings of the new system improved over time until some point

Robots in Time:
How User Experience in Human-Robot Interaction Changes over Time

Roland Buchner, Daniela Wurhofer, Astrid Weiss, and Manfred Tscheligi

HCI& Usability Unit, ICT&S Center, University of Salzburg, Austria
firstname.lastname@sbg.ac.at

Abstract. This paper describes a User Experience (UX) study on industrial robots in the context of a semiconductor factory cleanroom. We accompanied the deployment of a new robotic arm, without a safety fence, over one and a half years. Within our study, we explored if there is a UX difference between robots which have been used for more than 10 years within a safety fence (type A robot) and a newly deployed robot without fence (type B robot). Further, we investigated if the UX ratings change over time. The departments of interest were the oven (type A robots), the etching (type B robot), and the implantation department (type B robot). To observe experience changes over time, a UX questionnaire was developed and distributed to the operators at three defined points in time within these departments. The first survey was conducted one week after the deployment of robot B (n=23), the second survey was deployed six months later (n=21), and the third survey was distributed one and a half years later (n=23). Our results show an increasing positive UX towards the newly deployed robots with progressing time, which partly aligns with the UX ratings of the robots in safety fences. However, this effect seems to fade after one year. We further found that the UX ratings for all scales for the established robots were stable at all three points in time.

Keywords: Industrial Robots, Measurement, Semiconductor Factory, User Experience.

1 Introduction

For effective and highly productive industrial manufacturing, robots have already shown their usefulness in many sectors of production. With that kind of automation, a vast, cheap, and fast production has become reality. However, most of these systems are placed within a safety fence. During production, no human is allowed to enter the working space of the robot and therefore restricting access, any interaction, and/or cooperation with the robot. However, there are claims that more powerful human-robot interaction with the human and the robot working as a team is needed in order to be highly competitive [1]. That means it is necessary to break the general known paradigm of strictly separating

G. Herrmann et al. (Eds.): ICSR 2013, LNAI 8239, pp. 138–147, 2013.

UX Definitions

- "A consequence of a **user**'s internal state, the characteristics of the designed **system** and the **context** within which the interaction occurs" (Hassenzahl&Tractinsky 2006)*
- "All aspects of the **end-user**'s interaction with the company, its services, and its products" (Nielsen Norman Group)
- "The quality of experience a **person** has when interacting with a specific design" (Uxnet, online)

* Marc Hassenzahl & Noam Tractinsky (2006): User experience - a research agenda, Behaviour &Information Technology, 25:2, 91-97

UX list

satisfying helpful fun
enjoyable motivating provocative
engaging challenging surprising
pleasurable enhancing sociability rewarding
exciting supporting creativity emotionally fulfilling
entertaining cognitively stimulating

boring unpleasant
frustrating patronizing
making one feel guilty making one feel stupid
annoying cutesy
childish gimmicky

(Rogers, Sharp, Preece; Interaction design; 2011)

Values

The Three Laws of Robotics *(Isaac Asimov, I Robot)*

- A robot may not injure a human being or, through inaction, allow a human being to come to harm
- A robot must obey the orders given it by human beings except where such orders would conflict with the First Law
- A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws
- The Zeroth Law: A robot may not injure humanity, or, by inaction, allow humanity to come to harm

- What about environment, peace, justice…UN Sustainable Development Goals?

Trust


Harvard Business Review

Latest Magazine Popular Topics Podcasts Video Store The Big Idea Visual Library

What It Will Take for Us to Trust AI

by Guru Banavar

The early days of artificial intelligence have been met with some very public hand wringing. Well-respected technologists and business leaders have voiced their concerns over the (responsible) development of AI. And Hollywood's appetite for dystopian AI narratives appears to be bottomless.

This is not unusual, nor is it unreasonable. Change, technological or otherwise, always excites the imagination. And it often

- "If we are ever to reap the full spectrum of societal and industrial benefits from artificial intelligence, we will first need to trust it"

- Trust of AI systems will be earned over time
- One need to recognize and minimize bias, introduced for example by data sets as chatbot Tay who become racist by reading tweets
- Algorithms should be able to explain their suggestions or decisions
- *-> We'll learn more about trust in Human – AI partnership session*

Evaluation – takeaways

- Importance of the evaluation cannot be overestimated
- Consider the big picture – application domain, task, users, context of use
- Consider type of the system you are evaluating and interconnections between the data, algorithms and people
- Consider using existing design guidelines and using/extending existing instruments for measuring usability and UX
- Focus of the evaluation is moving towards values, trust, ethics

Your turn - Task 3 What to evaluate?

LUDVIG

https://www.youtube.com/watch?v=U9KrEcn4W3Q

Which UX dimensions would you evaluate?
Which values should be adresses by design?

Group work – 10 minutes discussion

References

- McCarthy J (2007) What is artificial intelligence. Technical report, Stanford University. http://www-formal.stanford.edu/jmc/whatisai.html
- Minsky ML (ed) (1968) Semantic information processing. MIT Press, Cambridge
- McCorduck P (2004) Machines who think. A K Peters/CRC Press, Boca Raton
- Hosanagar, A human's guide to machine intelligence, Viking, 2019
- HERNÁNDEZ-ORALLO, EVALUATION IN ARTIFICIAL INTELLIGENCE: FROM TASK-ORIENTED TO ABILITY-ORIENTED MEASUREMENT , J. ARTIF INTELL REV (2017) 48: 397.
- Paul Lukowicz and Philipp Slusallek. 2018. How to avoid an AI interaction singularity. Interactions 25, 5 (August 2018), 72-78.

- Buchner R., Wurhofer D., Weiss A., Tscheligi M. (2013) Robots in Time: How User Experience in Human-Robot Interaction Changes over Time. In: Herrmann G., Pearson M.J., Lenz A., Bremner P., Spiers A., Leonards U. (eds) Social Robotics. ICSR 2013. Lecture Notes in Computer Science, vol 8239. Springer, Cham
- Marc Hassenzahl & Noam Tractinsky (2006): User experience - a research agenda, Behaviour &Information Technology, 25:2, 91-97
- Banavar, What It Will Take for Us to Trust AI, Harvard Business Review, November 2016