# Human-AI Interaction

## Introduction
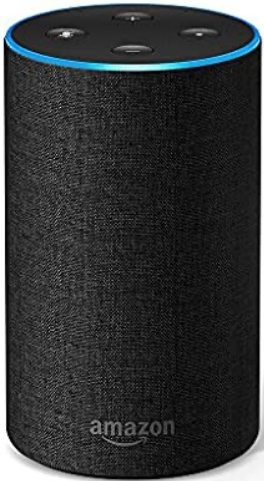
Luigi De Russis, Alberto Monge Roffarello

Politecnico di Torino

1859

e-Lite

# Let's Try to Set a Baseline...

- Classification? Clustering?

- Classification vs. regression?

- Unsupervised vs. supervised learning?

- Ontology?

- Cold start problem?

- Precision and recall?
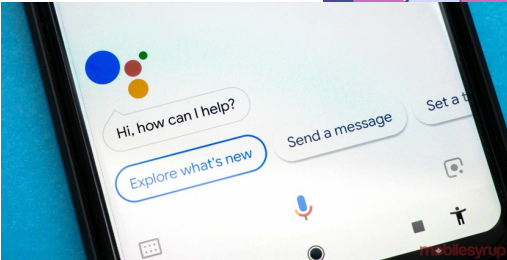
- Expert Systems?

# AI is everywhere!

# AI is everywhere!

- When it "works", it's *great!*

# AI is everywhere!

- When it "fails", it does it *spectacularly…*
  - e.g., Tesla Smart Summon, https://www.youtube.com/watch?v=VbVoTK-lM0o

**You**
what is 3*4+9*9? give the answer first and then the explanation

**ChatGPT**
The answer is 87.

Explanation:
$3 \times 4 + 9 \times 9$
$= 12 + 81$
$= 87$

# AI is everywhere!

- … and/or it is **<u>very</u>** problematic, for instance:

  o *"IBM boasted that its AI could 'outthink cancer.' Others say computer systems that read X-rays will make radiologists obsolete…"*

  o *"Systems developed in one hospital often flop when deployed in a different facility.*
  *Software used in the care of millions of Americans has been shown to discriminate against minorities.*
  *And AI systems sometimes learn to make predictions based on factors that have less to do with disease than the brand of MRI machine used, the time a blood test is taken or whether a patient was visited by a chaplain."*

  [source: https://www.scientificamerican.com/article/artificial-intelligence-is-rushing-into-patient-care-and-could-raise-risks/]

- Why?

# A Possible Reason: The Typical Approach
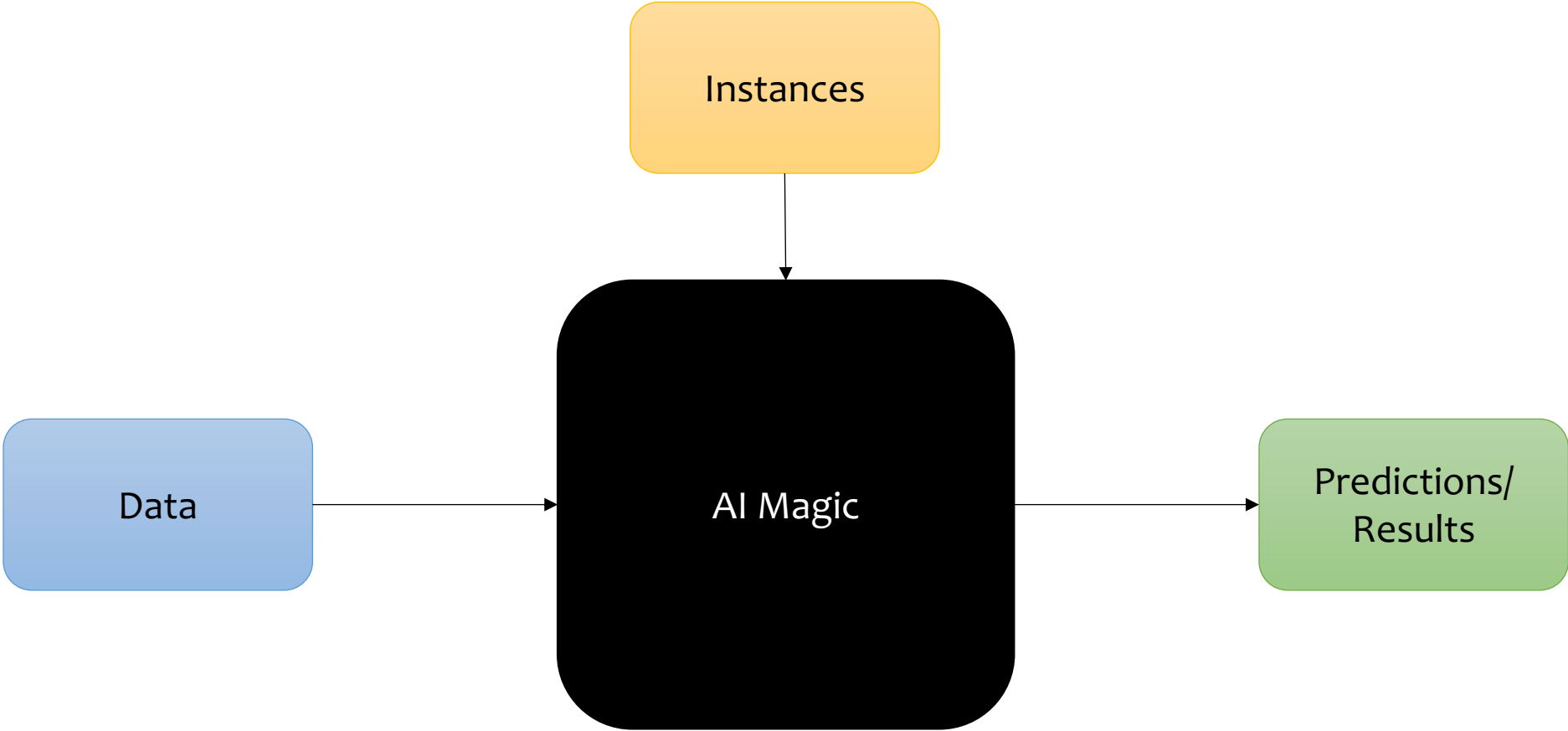
# Motivation

- *Most AI/ML courses consider "user interfaces" or humans as an* afterthought, near the end
  - several times they do not even think about "humans" ☹
  - they focus on algorithms/models, basically
- Why do not consider people from the *beginning,* and along the design, algorithmic choices, … in an *iterative* way?!

# Ultimately, AI Systems Are…

- Designed by *humans*

- To solve a problem framed by *humans*

- With *humans* taking specific choices (e.g., which algorithm to use)

- Evaluated and tested by *humans*

- With an outcome for *humans* (often)

- Presented to *humans* with a user interface (sometimes)

# Algorithms As The (Main) Answer?

- Algorithms are **not always** the "answer"
  - for instance: if you go to Netflix for the first time, what should it recommend you watch?
  - this is the *cold start problem*, and it is not really and fully solved
    - algorithmically speaking, at least

⇒ A **suitable** user interface is **critical** to overcome some limitations!

- Keeping people involved and considering them since the beginning is **fundamental**!

# Challenges

- How to ensure that people use AI-powered interfaces and systems with joy and trust rather than frustration and disappointment?

- How can we design and evaluate human-centered AI systems?

- How can we avoid (or minimize) problems, failures, ethical issues, … in AI systems?

# People & Computers

"The two hardest problems in computer science are: (i) people, (ii), convincing computer scientists that the hardest problem in computer science is people, and, (iii) off by one errors."

Prof. Jeffrey P. Bigham, 2018
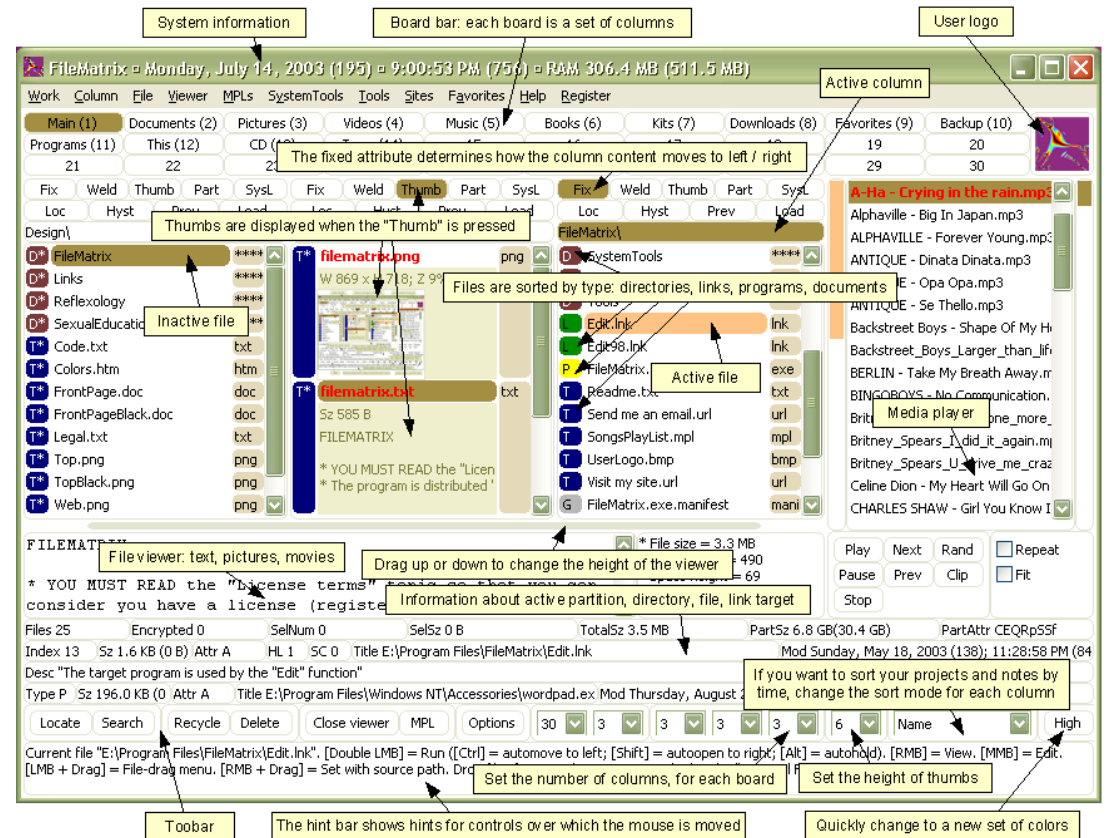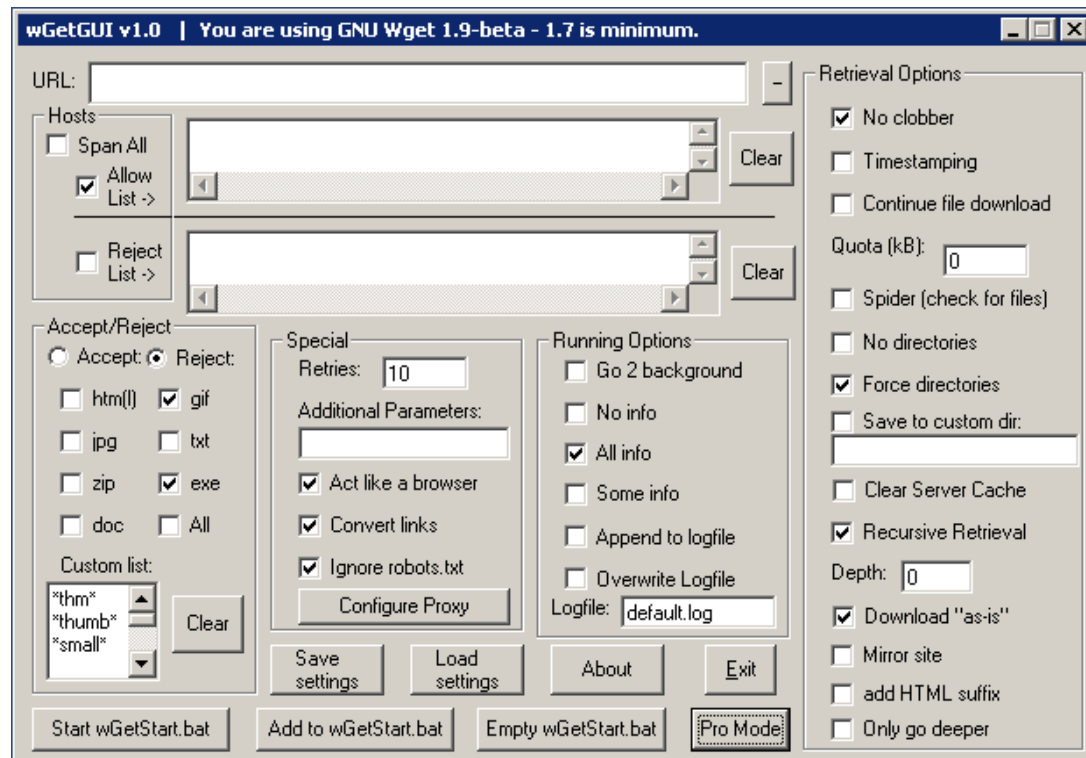
http://www.cs.cmu.edu/~jbigham/

# You Know… Developers…

# Human-AI Interaction: a Ph.D. Course

- *"Both [AI and HCI] explore the nexus of computing and intelligent behavior."*

  [source: Jonathan Grudin, "AI and HCI: Two Fields Divided by a Common Focus", 2009,
  https://doi.org/10.1609/aimag.v30i4.2271]

- We will build upon both disciplines.

What is *different* in interactive AI systems?

# What is Different in Interactive AI Systems?

- AI-based systems are typically performed under **uncertainty**
  - often producing false positives and false negatives
  - they can be **incorrect**
- They may demonstrate unpredictable behaviors that can be *disruptive, confusing, offensive,* and even *dangerous* for users

# What is Different in Interactive AI Systems?

- Before designing an AI system, we should ask:
  - what problems should be solved?
  - which AI approaches match human expectations given a problem?
  - which problems can be solved well enough for a particular use case?

# Human-AI Interaction: a Ph.D. Course

- Great interest in **research**!
  - Human-centered AI, eXplainable AI, …
  - you are "training" as researchers, after all

- The course will give some starting points and directions
  - research-based
  - if you want, you can go deep on different topics
  - general principles and ideas still apply

# (A Few) Other Relevant Courses @ PoliTo

- AI
  - Artificial Intelligence Safety, Ph.D. course
  - Machine Learning for pattern recognition, Ph.D. course
  - Mimetic Learning, Ph.D. course
  - Neuro-symbolic artificial intelligence, Ph.D. course

- HCI
  - Human-Computer Interaction, 6 credits, M.S. in Computer Engineering
  - Information visualization and visual analytics, Ph.D. course

# What Do I Mean For AI, here?

- Umbrella Term
  - Machine Learning, Knowledge Representation, Generative AI, …

- Various Application Areas
  - Computer Vision, Natural Language Understanding and Processing, …


- "Computers doing things that we expect people to be able to do"
  - Recognize if a photo contains a chair
  - Compute directions from here to Ikea
  - Infer that a chair is a piece of furniture
  - Recommend a movie

# What Do <u>You</u> Mean For AI?

- "Something that **thinks and act as a human** would (or in some cases as a **superhuman**)" (x2)

- "A machine with the human-like capabilities of **sensing, thinking and learning**" (x2)

- "A digital tool that **meets complex user's needs** reducing the necessary inputs"

- "A great tool to **help and enhance** human **work**"

- "A useful tool to go **beyond the limits** of many disciplines"

- "Providing **intelligence** to machines through mathematical models and their interaction"

- "Is the **encyclopedia** that came to life"

- "A **third brain hemisphere**"

- "An attempt at **emulating the potential of the** human **brain** made by humans, with all the **limits** of the human brain"

- "It's like **magic**, but with **algorithms!**" (x2)

- "Systems that **treat information automatically**"

# What Did <u>Your Colleagues</u> Mean For AI? (2022)

- "That one nerd-thing that is **thrilling** and **disturbing** at the same time" (x3)

- "Well-written algorithms that seem to be **intelligent**" (x3)

- "A clever organization of **deterministic operations** directed by a cleverer randomness"

- "An algorithm able to manage something **never seen before** (the opposite of being deterministic), also something able to pretend (very well) to **understand human behaviors**"

- "Machines doing stuff and **solving problems** on their own"

- "A bunch of math for **cool applications**"

- "**Smart** machines capable of performing **any tasks**"

- "Something able to **discover the unknown**"

- "An attempt to **replicate** how humans think" (x2)

- "An invention **as big as the fire** in the Stone Age", "A child **prodigy**"

- "A nice trick to **increase productivity and efficiency** in all walks life", "Tool to **enhance life**"

- "Like a human being but **better** (hoping **it doesn't kill us all**)"

# What Did <u>Your Colleagues</u> Mean For AI? (2020)

- "Computers **taking decisions** as they are 'thinking'" (x2)

- "A machine to **answer questions** in a reasonable way"

- "[A system] **self-conscious, explainable** and show **creative** behaviors"

- "A tool that emulates the capacity of the humans to **make decisions**" (x2)

- "Human **empowerment** through intelligent **data processing**", "**Supportive**

intelligence for the human being"

- "Algorithms great for **solving some/hard problems**" (x2)

- "**Transferring** human **intelligence** into machines"

- "Machines capable of **reasoning**"

- "A tool to let humans be humans by **delegating tasks** to machines"

# Course Logistics

# "Teaching Philosophy"

- Put persons first!
  - different backgrounds and expectations in this room
  - how to do something "good" for all of you?

- Interactivity

- Learn by doing, do by learning
  - mix of lectures, practical exercises, and readings

- To learn something, teach it
  - Panel and workshop-style sessions

# About You (hello!)

- 35 (enrolled) students
  - 17 from the Ph.D. in Computer and Control Engineering
  - 3 from the Ph.D. in Electrical, Electronics and Communications Engineering
  - 3 from the Ph.D. in Architecture, History and Project
  - 3 from the Ph.D. in Bioengineering and Medical-Surgical Sciences
  - 2 from the Ph.D. in Energetics
  - 2 from the Ph.D. in Management and Production Engineering
  - 1 from the National Ph.D. in Artificial Intelligence
  - 3 from the Ph.D. in Civil and Environmental Engineering
  - 1 from the Ph.D. in Urban and Regional Development

# About You (hello!)

- Different (research) interests
    - Art + Science
    - AR and MR
    - Digitalization for products
    - Computational mechanics
    - Predictability of digital plans on humans
    - Intracoronary physiology and cardiology
    - Vehicle-To-Everything cooperation
    - Medical Image Segmentation and DL
    - LLMs meet databases
    - Gamification + Software Engineering
    - Neural Networks for Hardware
    - AI + Architecture
    - Explainable, Responsible AI
    - …

# About Us

- Luigi De Russis
  - Associate Professor
  - Department of Control and Computer Engineering
  - Research Topic: HCI in Complex Settings
  - [luigi.derussis@polito.it](mailto:luigi.derussis@polito.it)

- Alberto Monge Roffarello
  - Assistant Professor (RTDb)
  - Department of Control and Computer Engineering
  - Research Topic: Digital Wellbeing
  - [alberto.monge@polito.it](mailto:alberto.monge@polito.it)

# Course Topics

- Introduction to Human-AI Interaction

- Trade-offs and perspectives in Human-AI Interaction
  - Augmenting or replacing people?
  - Direct manipulation or agents?

- Designing and evaluating human-centered AI systems
  - Guidelines and methods
  - Data, bias, and trust

- Paradigms for Human-AI Interaction
  - Smart interfaces and conversational agents

- Hands-on sessions: design and prototyping a conversational agent

# Course Information

- Material
  - https://elite.polito.it ->Teaching -> Human-AI Interaction (01UJUIU)
  - short link: https://bit.ly/polito-haii
  - Slides, exercises, videos, etc.
- Students are encouraged to attend the classes with their laptops, to work on the proposed exercises

| Date | Time | Room | Type | Topic | Video | Teacher |
|---|---|---|---|---|---|---|
| **Class #1** | | | | | | |
| 22/01/2024 | 09:00-11:30 | 8I | Lecture | Introduction and fundamentals | | Luigi De Russis |
| 22/01/2024 | 11:30-13:00 | 8I | Lecture | Madness session<br><br>Reading panels: instructions | | Luigi De Russis |
| **Class #2** | | | | | | |
| 26/01/2024 | 09:00-11:00 | 8I | Exercise | Reading panels (to be *prepared* beforehand) | | Luigi De Russis |
| 26/01/2024 | 11:00-13:00 | 8I | Lecture | Perspectives on Human–AI Interaction | | Luigi De Russis |
| **Class #3** | | | | | | |
| 31/01/2024 | 09:00-11:00 | 8I | Lecture | Designing and evaluating interactive AI systems | | Alberto Monge Roffarello |
| 31/01/2024 | 11:00-13:00 | 8I | Exercise | Design & evaluation workshop | | Alberto Monge Roffarello |
| **Class #4** | | | | | | |

# The Plan: Overview

- 5 classes
  - around 50% interactive lectures and 50% exercises

- 4 hours per class

- Schedule
  1. 22/01/2024 h. 9:00-13:00, room 8i
  2. 26/01/2024 h. 9:00-13:00, room 8i
  3. 31/01/2024 h. 9:00-13:00, room 8i
  4. 06/02/2024 h. 9:00-13:00, room 8i
  5. 12/02/2024 h. 9:00-13:00, room 8i

# The (Tentative) Detailed Plan

| Class | Type | Topic | Teacher(s) |
|---|---|---|---|
| 1 | Lecture | Course introduction, logistics, introduction to Human-AI Interaction | Luigi |
| 2 | Exercise | **Reading Panels.** Finalize group formation | Luigi |
| | Lecture | Perspectives on Human-AI Interaction | |
| 3 | Lecture | Designing and Evaluating Interactive AI Systems | Alberto |
| | Exercise | **Design & Evaluation Workshop** | |
| 4 | Lecture | Conversational Agents | Alberto |
| | Exercise | **Case Study: designing and implementing a conversational assistant** | |
| 5 | Exercise | **Case Study (cont'd).** | Alberto |
| | Exercise | **Case Study (cont'd). Final Presentation.** | Alberto, Luigi |

# Exam

**Three** practical activities, to be carried out **in class.**

1. Readings Panels (individual, <u>next week</u>) -> *to be prepared before the class!*

2. Design and Evaluation Workshop (in group, class #3)

3. Case Study Prototype and Presentation (in group, classes #4-5)


To **pass** the exam:
- o two activities completed with success – one *being* the case study
- o **MERIT** with all three activities (successfully) done

# Our Times...

- The preferred way to follow the course is *in person.*

- However, "life happens":
  - Lectures will be video-recorded and shared after each class (YouTube and Portale).
  - Group exercises can have *hybrid* groups, with <u>at least</u> one person of the group in the room.
  - Try hard to be in the class for the individual exercise (i.e., the panels). Send me an email if you **cannot really** be there <u>but</u> you want to join remotely.

# TODO: Group Composition

- Form a group for the upcoming activities

- **3-5 people** per group

- Fill up this spreadsheet:
    - [https://docs.google.com/spreadsheets/d/1KParqVgJ0YcU6mXLN5xGu21wLLf21L2YnaWUra1DolQ/](https://docs.google.com/spreadsheets/d/1KParqVgJ0YcU6mXLN5xGu21wLLf21L2YnaWUra1DolQ/)

- **Deadline:** January 30, 2024

# About Programming…

- Do you know "enough" programming?

- You need to know some Python (preferably)
  - other languages may be ok (e.g., JavaScript, Java, …)


- Needed for the **case study**, only
  - We will provide examples and projects to get started with (*~1*)
  - We will be here (*obviously!*)

# Questions?

# I Have Some Questions For You…

- I am a ML expert, a smart home enthusiast, and I applied AI in my home

- After an *adequate* period of data collection about my habits at home, I wrote a ML system to automatize my most frequent habits

- For instance, the AI detected that:
  o almost every morning, Mon-Fri, I wake up at 6:30
  o then, I turn on the light
  o I open the window for around 10 minutes
  o I start my coffee machine

- The system automatically executes these steps

# I Have Some Questions For You…

- Is it a good problem to solve?

- Does it solve the "morning routine" totally?

- What can go wrong?

- Any failures and possibility to recover?

- Better ways to do this?

- …

# Fundamentals

Human-Computer Interaction and Human-Centered AI

# AI+HCI

- "From the earliest times in the development of computers, activities in the field of Human-Computer Interaction (HCI), and Artificial Intelligence (AI) have been intertwined.  But as subfields of Computer Science, AI and HCI have always had a love-hate relationship."

- "Together, the community can make user interfaces a little less stupid and frustrating than they are today."

[source: Henry Lieberman, "User Interface Goals, AI Opportunities", AI Magazine, 2009, https://doi.org/10.1609/aimag.v30i4.2266]

# Human-Computer Interaction (HCI)



- A multi-disciplinary field

- Concerned with the **design, evaluation,** and **implementation** of interactive computing systems for human use
  - and with the study of **major phenomena** surrounding them



-  Involves two *entities* (the human and the computer) that determine each other behavior over time
  - framed in terms of humans' **goals** and related **tasks/pursuits**
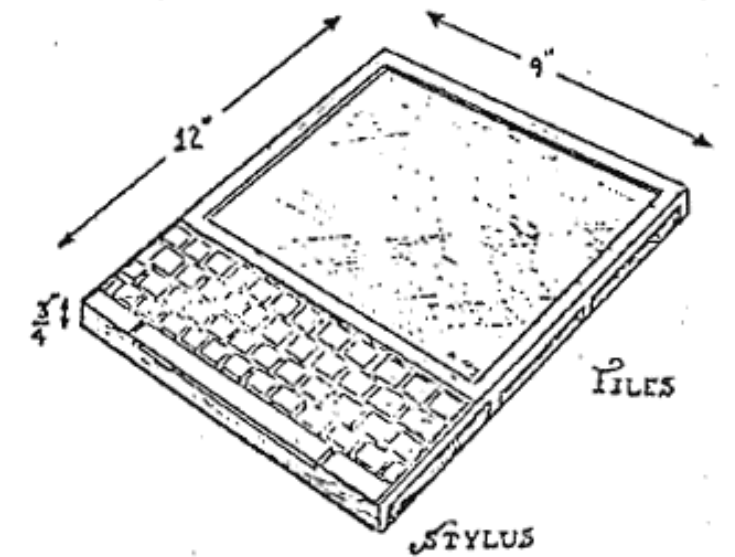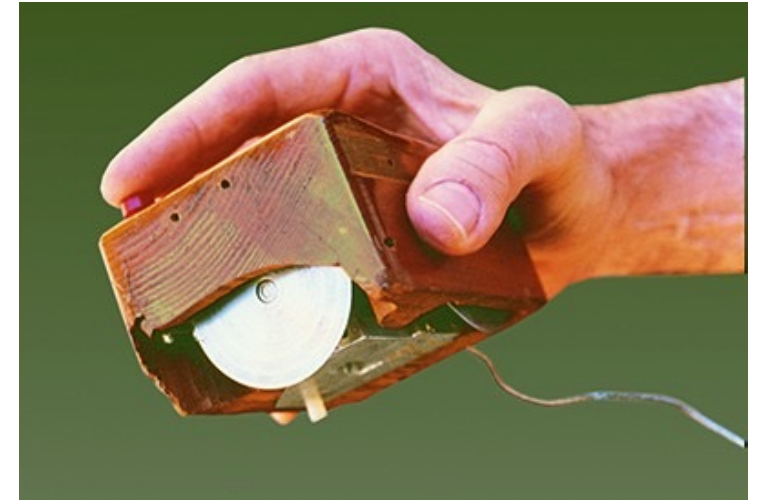
# What Is "Interaction" (in HCI)?

- Interaction is...

| Concept | View of interaction | Key phenomena and constructs | Good interaction | Example support for evaluation and design |
|---|---|---|---|---|
| Dialogue | a cyclic process of communication acts and their interpretations | mappings between UI and intentions; feedback from the UI; turn taking | understandable; simple, natural; direct | methods/concepts for guessability, feedback, mapping; walkthroughs |
| Transmission | a sender sending a message over a noisy channel | messages (bits); sender and receiver; noisy channels | maximum throughput of information | metrics and models of user performance |
| Tool use | a human that uses tools to manipulate and act in the world | mediation by tools; directness of acting in the world; activity as a unit of analysis | useful and transparent tools; amplification of human capabilities | compatibility in instrumental interaction; break down analysis |
| Optimal behavior | adapting behavior to goals, task, UI, and capabilities | rationality; constraints; preferences; utility; strategies | improves or reaches maximum or satisfactory utility | models of choice, foraging, and adaptation |
| Embodiment | acting and being in situations of a material and social world | intentionality; context; coupling | provides resources for and supports fluent participation in the world | studies in the wild; thick description |
| Experience | an ongoing stream of expectations, feelings, memories | non-utilitarian quality; expectations; emotion | satisfies psychological needs; motivating | metrics of user experience; experience design methods |
| Control | interactive minimization of error against some reference | feedforward; feedback; reference; system; dynamics | rapid and stable convergence to target state | executable simulations of interactive control tasks |

**Taken from:** Kasper Hornbæk & Antti Oulasvirta, What Is Interaction? In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*
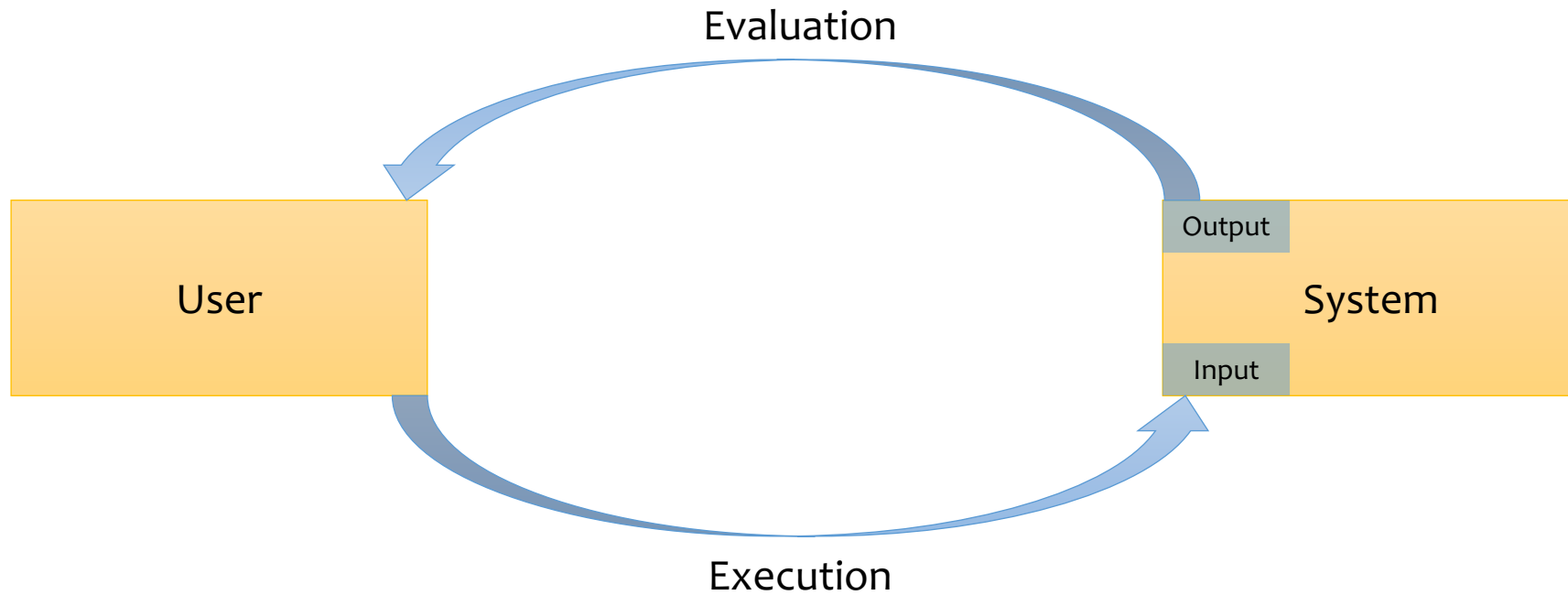
# What Is "Interaction" (in HCI)?

- Interaction…
  - **is not** the idea promoted and repeated in folk notions that a computer and a human are engaged
  - it concerns two entities – humans and computers – that determine each other's behavior over time
    - Their mutual determination can be of many types, including statistical, mechanical, and structural

- **Users**, with their **goals** and **pursuits**, are the ultimate metric of interaction
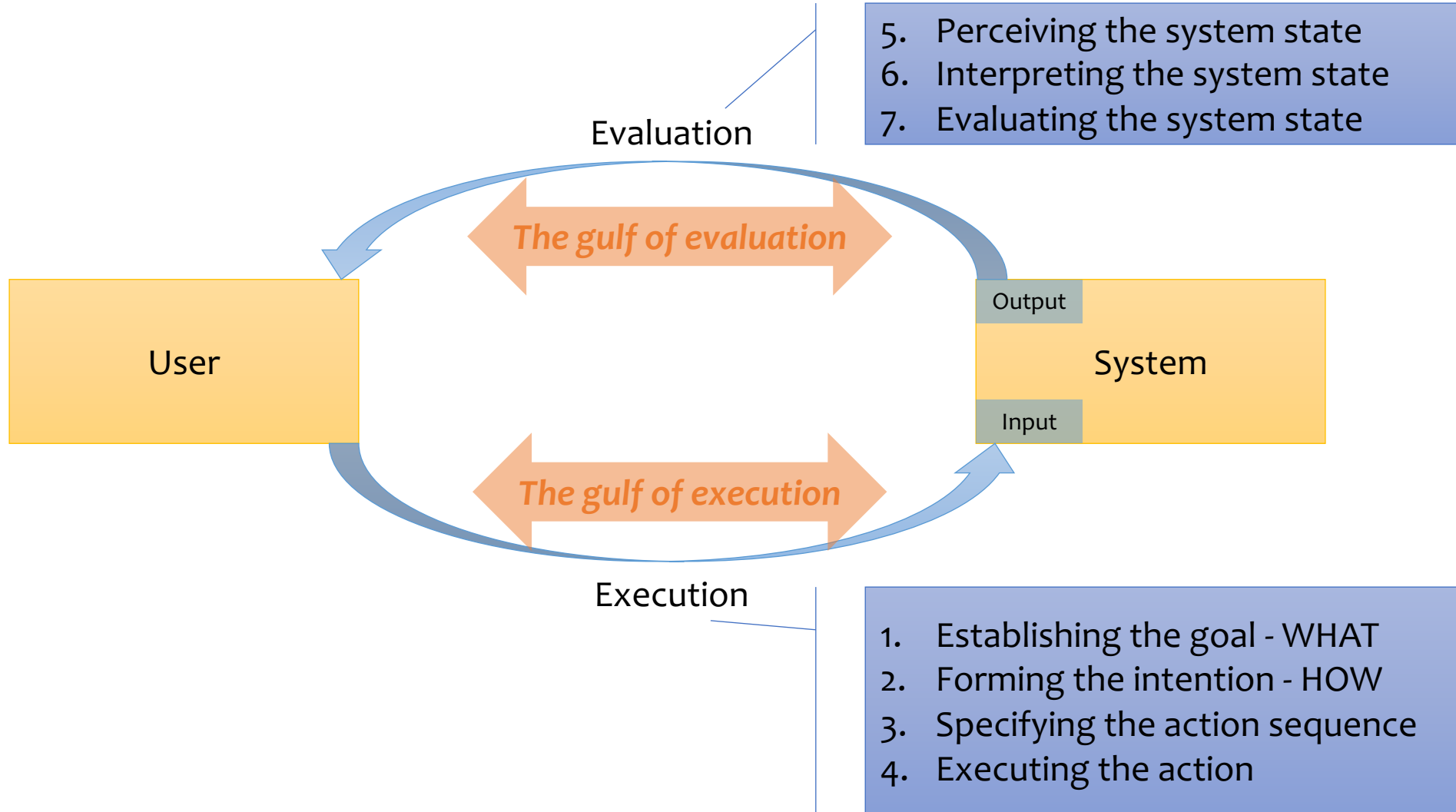
# Assumptions

- The **user** wants to accomplish some **goals,** in a specific application **domain**
  - Each domain has a specific jargon, set of possible processes and goals, artifacts and building blocks, ...

- **Tasks** are operations to manipulate the concepts of a domain
  - The goal is attained by performing one or more tasks

- Interaction studies the relation between User and System
  - The system possesses a **state** and "speaks" a **core language**
  - The user possesses a **state,** that includes an **understanding** of the system's state, some **intention** to perform a task, and "speaks" the **task language**

# Norman's Model of Interaction

# Norman's Model of Interaction

Evaluation

5. Perceiving the system state
6. Interpreting the system state
7. Evaluating the system state

*The gulf of evaluation*

User

Output

System

Input

*The gulf of execution*

Execution

1. Establishing the goal - WHAT
2. Forming the intention - HOW
3. Specifying the action sequence
4. Executing the action

# A Matter Of Perspective

**Technology as an end**

**Technology as a mean to an end**

- The end: satisfying users' needs, attitudes, and expectations
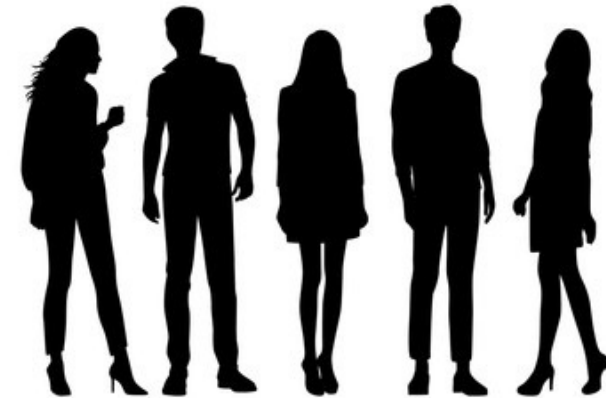


**People first and at the center!**

# Main Concepts

- **Key Attributes**
  - Usefulness: to accomplish what is required/expected
  - Usability: "do it easily... don't let me think"
  - Performance and robustness
  - Attractiveness and engagement

- **User Studies**
  - Understanding people and their needs
  - Analyzing their behavior (with technology)
  - Understanding how the designed solution affect people' perceptions, attitudes, and judgements

# A Human-Centered Approach

- **Early focus on people and tasks:** observing humans doing their tasks and then involving them in the design process

- **User-based evaluation:** users' reactions and performance to scenarios, simulations, and later to prototypes are observed, recorded, and analyzed

- **Iterative design:** when problems are identified through user testing, fix them and carry out more tests

**Benefits: ROI, Safety, Ethics, Innovation, …**

# Human-Centered AI

- Human-centered AI focuses on **amplifying, augmenting,** and **enhancing** human performance in ways that make AI systems **reliable, safe,** and **trustworthy**

- Shift from measuring **only** algorithm performance to evaluating human performance and satisfaction, with **human-centered** and participatory approaches (for evaluation, too)

Ben Shneiderman, *Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems.* ACM Transactions on Interactive Intelligent Systems, Vol. 10, No. 4, Article 26, 2020, https://doi.org/10.1145/3419764

# A Paradigmatic Change
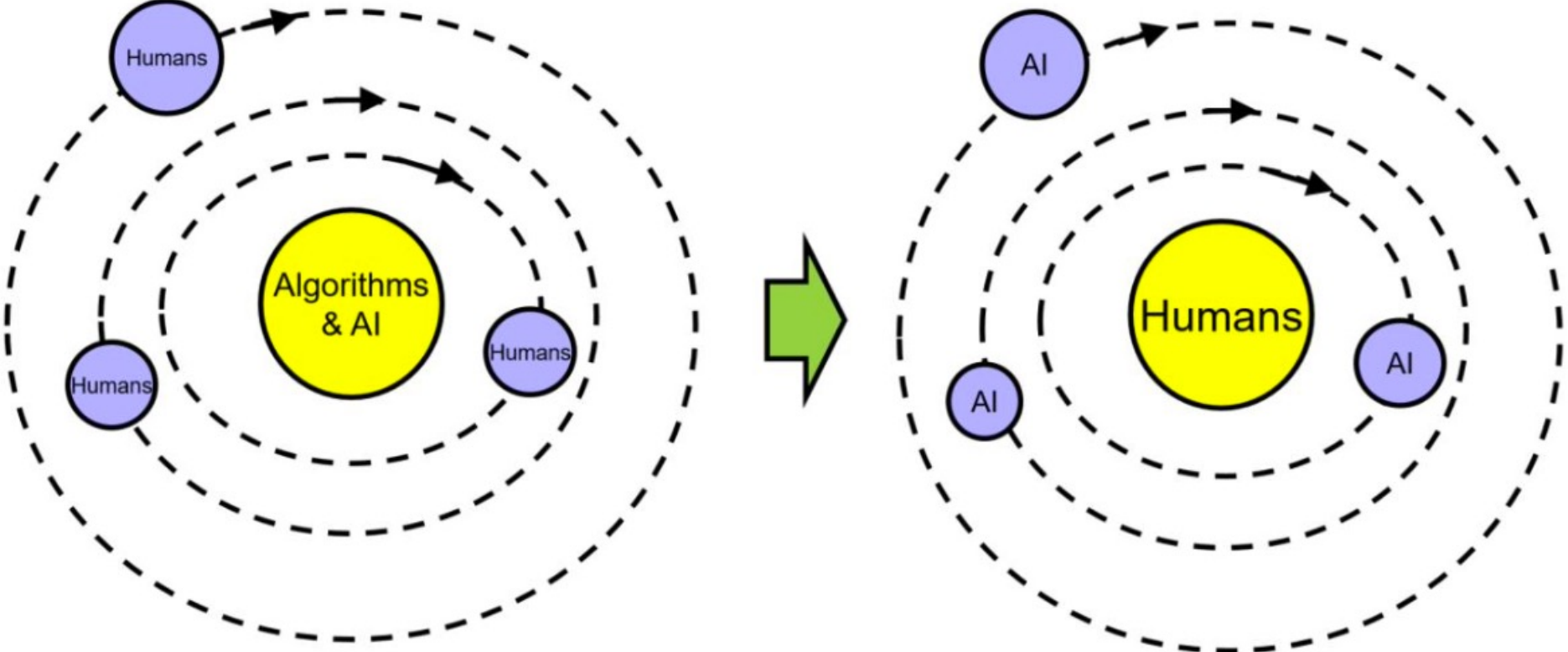
# License

- These slides are distributed under a Creative Commons license "**Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)**"

- **You are free to:**
  - **Share** — copy and redistribute the material in any medium or format
  - **Adapt** — remix, transform, and build upon the material
  - The licensor cannot revoke these freedoms as long as you follow the license terms.

- **Under the following terms:**
  - **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
  - **NonCommercial** — You may not use the material for commercial purposes.
  - **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
  - **No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

- https://creativecommons.org/licenses/by-nc-sa/4.0/