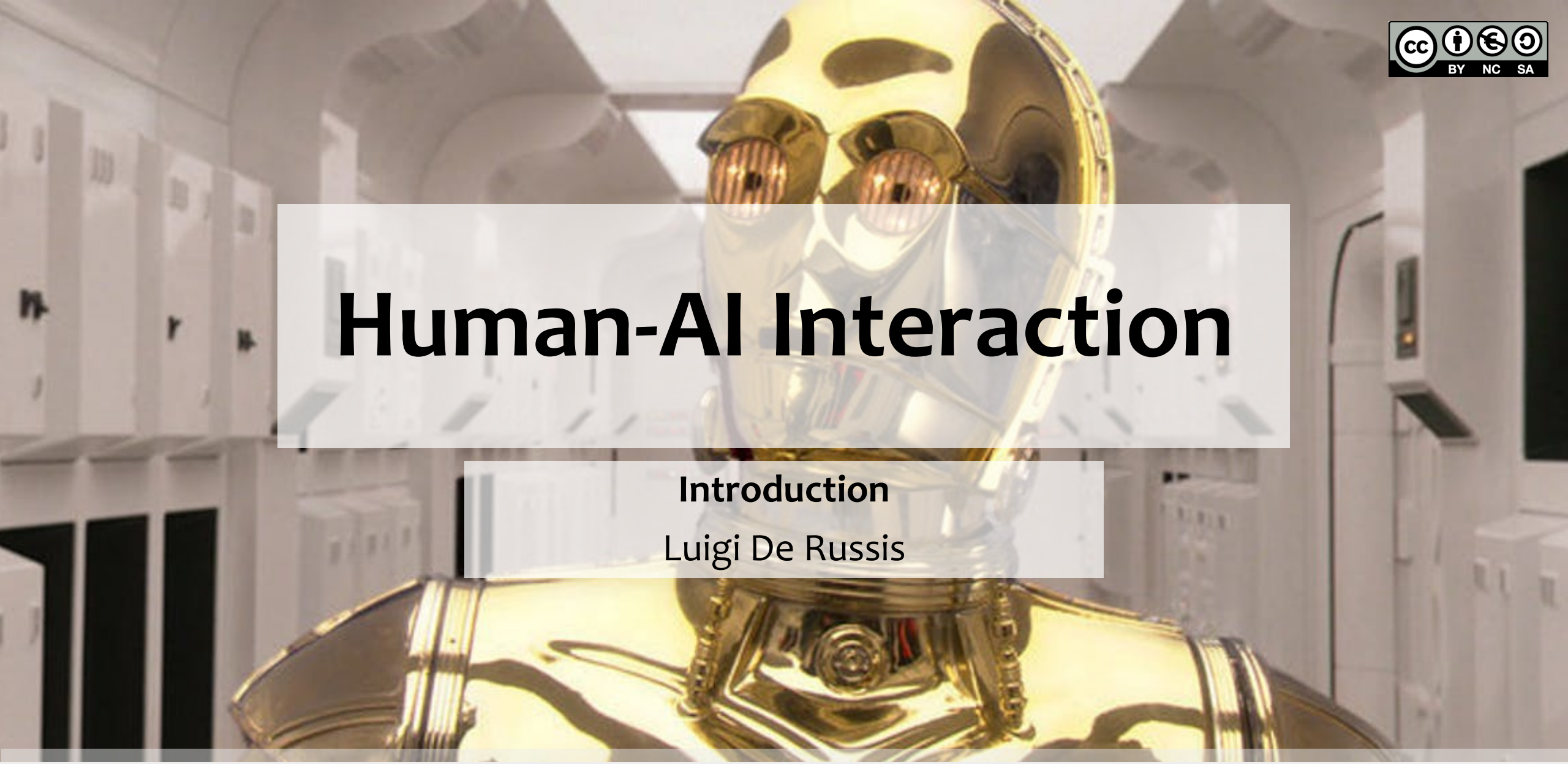


Human-AI Interaction

Introduction

Luigi De Russis

POLITECNICO DI TORINO

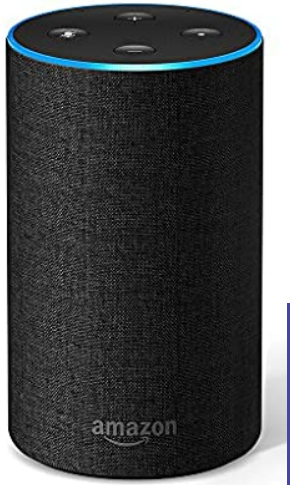
Academic Year 2019/2020

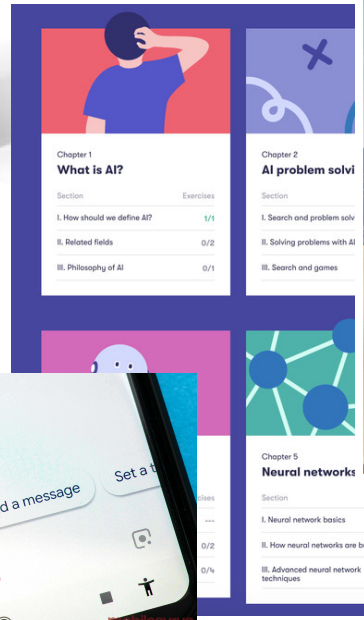
e-Lite

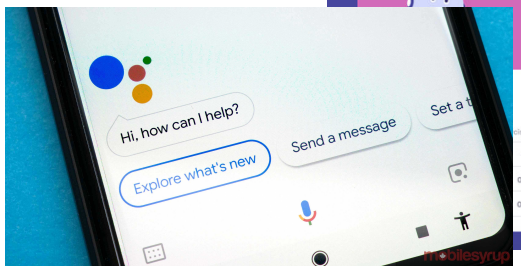
Let's Try to Set a Baseline…

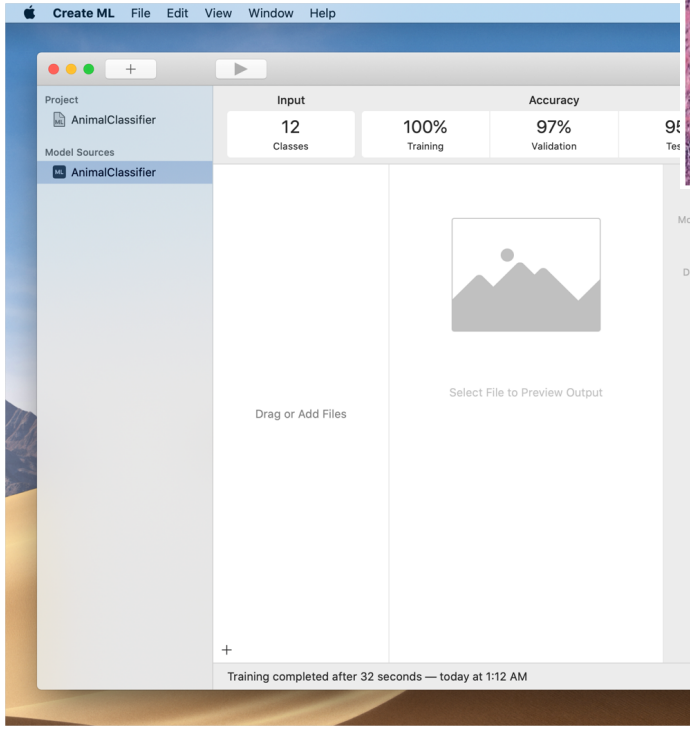
- Classification? Clustering?
- Classification vs. regression?
- Unsupervised vs. supervised learning?
- Ontology?
- Cold start problem?
- Precision and recall?
- Expert Systems?

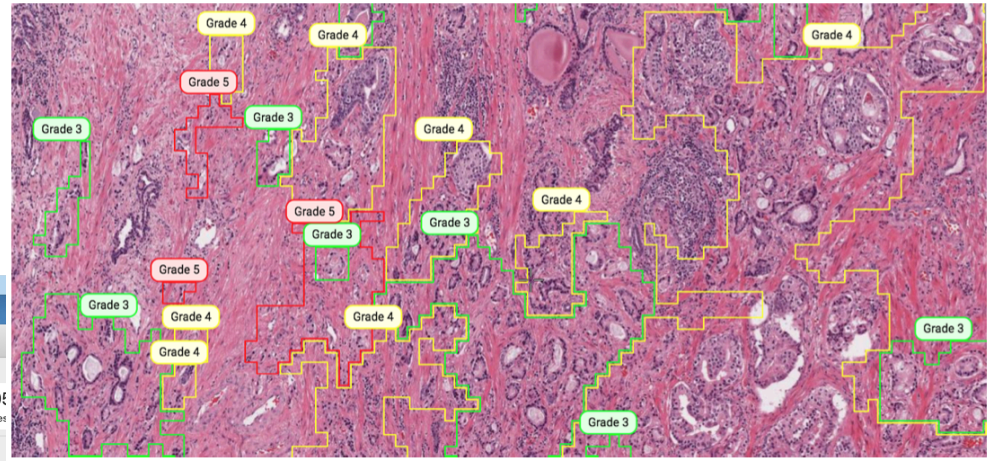
AI is everywhere!











AI is everywhere!

- When it "works", it's *great*!
- When it "fails", it does it *spectacularly*…
  - Tesla Smart Summon, https://www.youtube.com/watch?v=VbVoTK-lM00
  - Alexa, https://www.youtube.com/watch?v=QFpUN3kYTDA

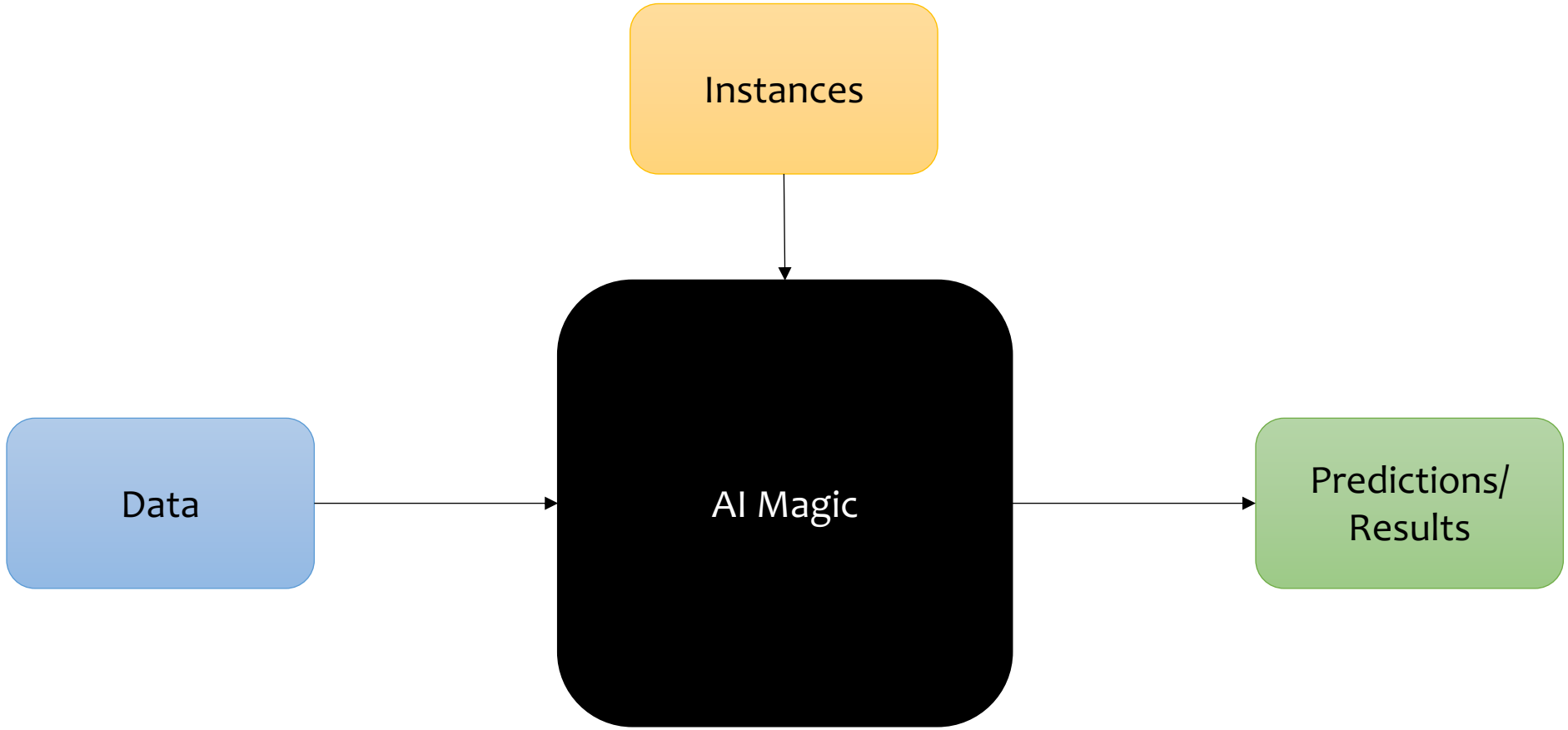
AI is everywhere!

- … and/or it is ***very*** problematic, e.g.,
  - *"IBM boasted that its AI could 'outthink cancer.' Others say computer systems that read X-rays will make radiologists obsolete…"*
  - *"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

Motivation

- 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 **in-the-loop** and considering them since the beginning is **fundamental**!

Challenges

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

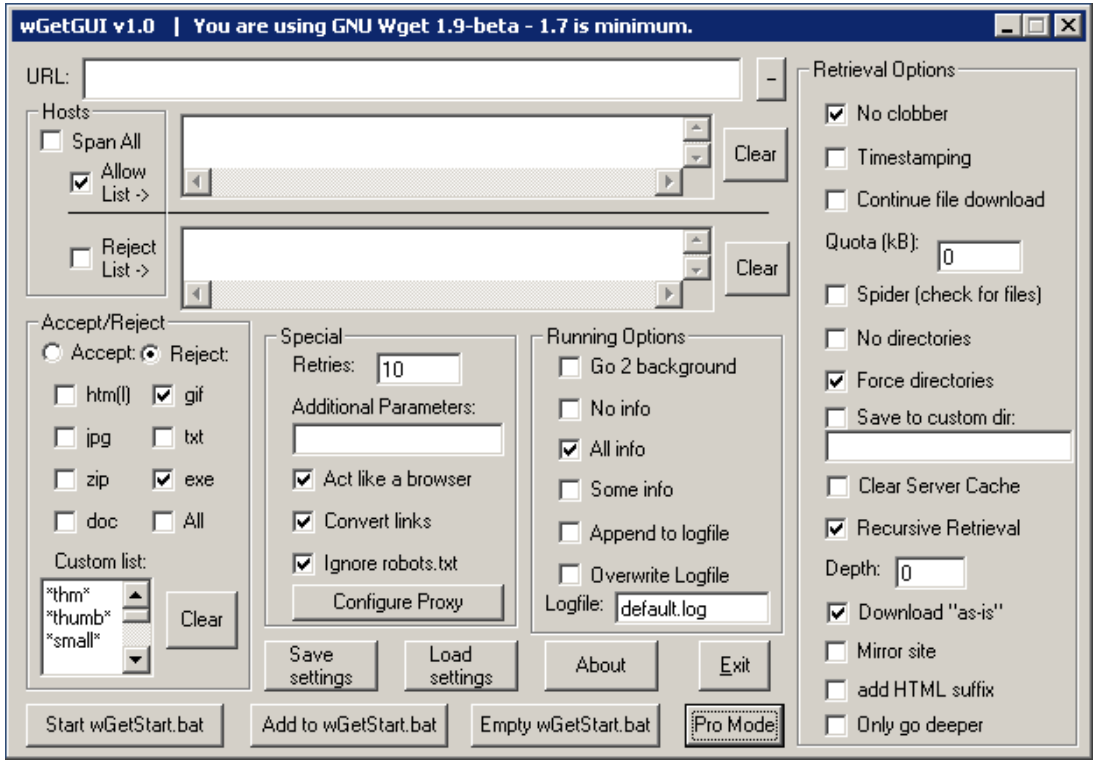
"The two hardest problems in computer science are: (i) people, (ii), convincing computer scientists that the hardest problem in computer science is people, and, (iii) off by one errors."

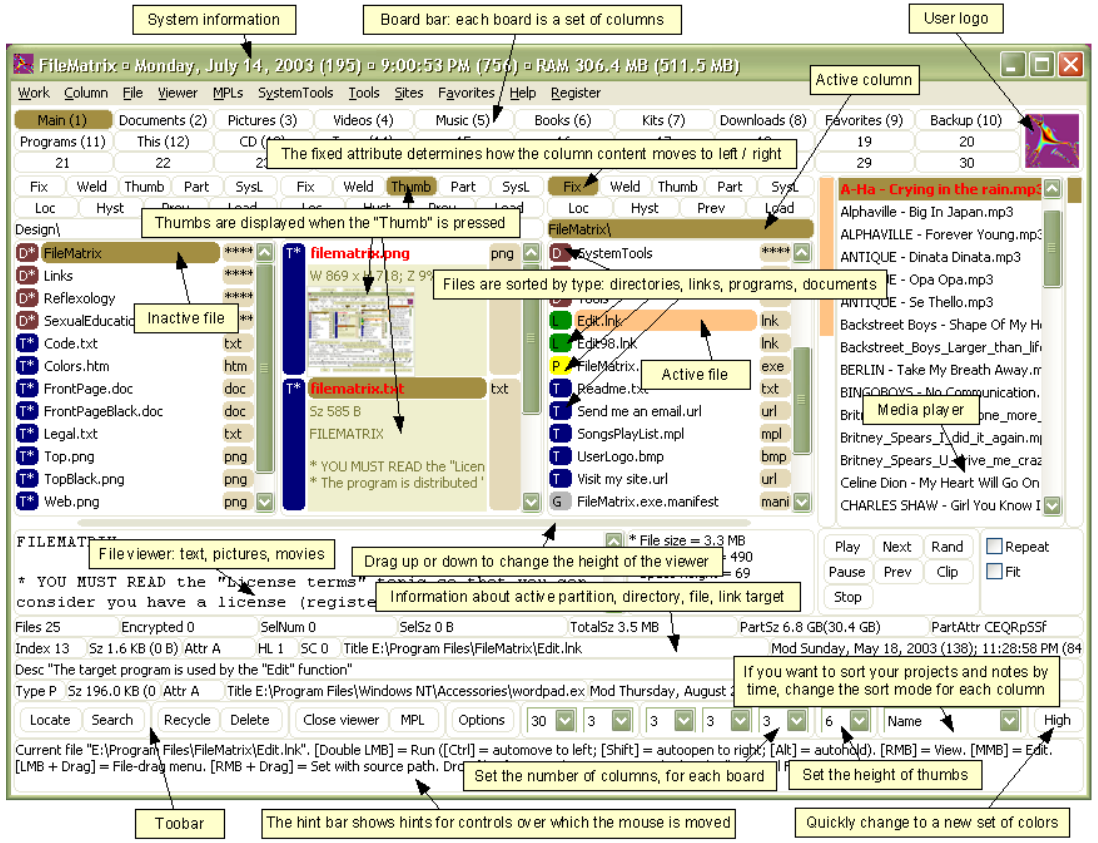
Prof. Jeffrey P. Bigham, 2018

http://www.cs.cmu.edu/~jbigham/

"People are a mess"

You Know… Developers…





https://thedailywtf.com/articles/Classic_WTF_-_Enter_the_Matrix

Human-AI Interaction: a Ph.D. Course

- At the intersection of AI and HCI
  - "*Both [AI and HCI] explore the nexus of computing and intelligent behavior.*"
    - 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
- There is something *different* about building interactive systems that include AI: AI is uncertain and may be incorrect
  - 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**!
  - 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

Some Other Relevant Courses @ PoliTo

- AI
  - Machine Learning and Artificial Intelligence, 6 credits, M.S. in Computer Engineering
  - Mimetic Learning, Ph.D. course
  - Semantic Web, Ph.D. course (2020/2021)
- HCI
  - Human-Computer Interaction, elective course, 6 credits, M.S. in Computer Engineering
  - Human-Machine Interaction, Ph.D. course

What Do I Mean For AI, here?

- Umbrella Term
  - Machine Learning, Knowledge Representation, Evolutionary Algorithms, …
- 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 You Mean For AI?

- "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 Contents

My "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
    - programming included!
- To learn something, teach it
  - Panel and workshop-style sessions

About You (hello!)

- 38 (enrolled) students
  - 18 from the Ph.D. in Computer and Control Engineering
  - 11 from the Ph.D. in Electrical, Electronics and Communications Engineering
  - 4 from the Ph.D. in Management, Product, and Design
  - 2 from the Ph.D. in Mechanical Engineering
  - 1 from the Ph.D. in Aerospace Engineering
  - 1 from the Ph.D. in Pure and Applied Mathematics (UniTo + PoliTo)
  - 1 from the Ph.D. in Computer Science (UniTo)
- Different (research) interests
  - Autonomous Vehicles, NLP, Computer Vision, Multibody Dynamics, Data Ethics and Quality, Mixed Reality, Health, …

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, explainability, and trust
- Case study on conversational agents and chatbots
  - Hands-on sessions: design and prototyping

Course Information

- Material
  - https://elite.polito.it ->Teaching -> Current Courses -> 01UJUIU – Human-AI Interaction
  - short link: http://bit.ly/polito-humanai
  - Slides, exercises, readings, etc.
- How to contact me
  - luigi.derussis@polito.it
  - Department of Control and Computer Engineering
- Students are encouraged to attend the classes with their laptops, to work on the proposed exercises

The Plan: Overview

- 5 classes
- 4 hours per class
  - 2 hours -> lecture
  - 2 hours -> practical activities
  - with a break, in-between!
- Schedule
  1. 16/01/2020 h. 14:30-18:30, aula 29B -> *we will try to finish at 18:00!*
  2. 23/01/2020 h. 14:00-18:00, aula 5N
  3. 30/01/2020 h. 14:00-18:00, aula 5N
  4. 05/02/2020 h. 9:00-13:00, aula 9S
  5. 12/02/2020 h. 9:00-13:00, aula 5N

The (Tentative) Detailed Plan

Week	Type	Topic
1	L	Course introduction, logistics, introduction to HAII.
	E	Madness session. AI in the World: Journey Map.
2	L	Perspectives on HAII.
	E	**Readings Panel.**
3	L	Designing and Evaluating HAII Systems.
	E	**Design & Evaluation workshop.**
4	L	Explainable AI. Conversational Assistants.
	E	**Case Study: Building a Conversational Assistant.**
5	E	**Case Study (cont'd).**
	E	**Case Study (cont'd): Final Presentation.**

Exam

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

1. Readings Panel (next week) -> *to be prepared before the class!*
2. Design and Evaluation Workshop (W3)
3. Case Study Prototype and Presentation (W4-5)

To **pass** the exam:

- 2 activities completed with success
- **MERIT** with all three activities (successfully) done

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
  - I will provide examples and projects to get started with (*>=1*)
  - I 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:
  - almost every morning, Mon-Fri, I wake up at 6:30
  - then, I turn on the light
  - I open the window for around 10 minutes
  - 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?
- …

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/