

Aspetos Profissionais e Sociais da Engenharia Informática

Talking about ML...
maybe you should understand what it means...

Rui L Aguiar, UA/IT

1




Last lesson....

- Standards
- Open source models, trabalhos derivados
- Marcas, IPR
- cybersegurança
- GPDR
- Engenharia informática
- Ecossistema

2

2

1



Today....

- AI – what is this
- Technical aspects
- Social and legal issues
 - Training, Bias and Security
 - Regulations
- Up next
 - AI rationale
 - The challenging future of AI
 - Scaling aspects
 - Hyperscalers


3

3



WHAT IS AI (ML?)


4



What is Artificial Intelligence ?

- the automation of activities we associate with human thinking, like decision making, learning ... ?
- the art of creating machines that perform functions that require intelligence when performed by people ?
- making computers that think?
- a field of study that seeks to explain and emulate intelligent behaviour in terms of computational processes ?
- the study of mental faculties through the use of computational models ?
- a branch of computer science that is concerned with the automation of intelligent behaviour ?

5




What is AI? *(class built)*

- Today?
 - A machine that seems to act as a human being in some tasks
 - A machine that does calculations and decisions that a human would not be able to do as fast
 - A machine that does tasks autonomously
 - A process that behaves as a human being would behave.
- Tomorrow?
 - Robotic presences (e.g. Wall-e)
 - Potential psicopath behaviour (simulate/lacking emotions)
 - cyborgs

6

6


3



Artificial Intelligence

- Artificial
 - Produced by human art or effort, rather than originating naturally.
- Intelligence
 - is the ability to acquire knowledge and use it" [Pigford and Baur]
- **So AI can be defined as:**
 - AI is the study of ideas that enable computers to be intelligent.
 - AI is the part of computer science concerned with design of computer systems that exhibit human intelligence(From the Concise Oxford Dictionary)


7



AI Multiple Definitions/Scopes

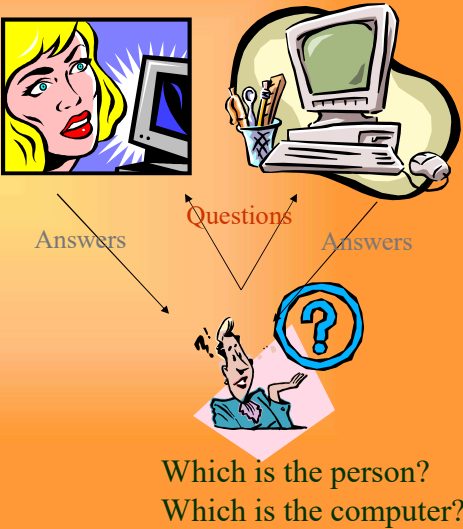
<ul style="list-style-type: none">■ The study of how to make programs/computers do things that people do better■ The study of how to make computers solve problems which require knowledge and intelligence■ The effort to make computers think ... machines with minds■ The automation of activities that we associate with human thinking (e.g., decision-making, learning...)	Thinking machines or machine intelligence
<ul style="list-style-type: none">■ The art of creating machines that perform functions that require intelligence when performed by people■ The study of mental faculties through the use of computational models■ A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes■ The branch of computer science that is concerned with the automation of intelligent behavior	Studying cognitive faculties

8




Review: The Turing Test

- 1950 – Alan Turing devised the Imitation Game
 - Ask questions of two entities, receive answers from both
 - If you can't tell which of the entities is human and which is a computer program, then you are fooled and we should therefore consider the computer to be intelligent

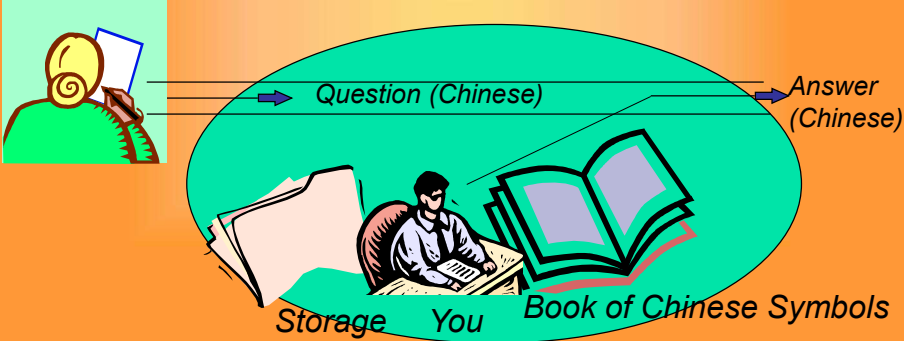


9




The Chinese Room Problem

- John Searle, to demonstrate that computers cannot be intelligent
 - The room consists of you, a book, a storage area (optional), and a mechanism for moving information to and from the room to the outside
 - a Chinese speaking individual provides a question for you in writing
 - you are able to find a matching set of symbols in the book (and storage) and write a response, also in Chinese




10



Searle's argument

- You were able to solve the problem of communicating with the person/user \Rightarrow you/the room passes the Turing Test
- But did you understand the Chinese messages being communicated?
 - since you do not speak Chinese, you did not understand the symbols in the question, the answer, or the storage
 - can we say that you actually *used* any intelligence?
- By analogy, since you did not understand the symbols that you interacted with, neither does the computer understand the symbols that it interacts with (input, output, program code, data)
- Searle concludes that the computer is not intelligent, it has no "semantics," but instead is merely a symbol manipulating device
 - the computer operates solely on syntax, not semantics
- He defines two categories of AI:
 - **strong AI** – the pursuit of machine intelligence
 - **weak AI** – the pursuit of machines solving problems in an intelligent way

11



Computers do Solve Problems

- Computers solve problems in a seemingly intelligent way
 - Where is the intelligence *coming* from?
- Different views against Searle's argument
 - The System's Response:
 - the hardware by itself is not intelligent, but a combination of the hardware, software and storage is intelligent
 - in a similar vein, we might say that a human brain that has had no opportunity to learn anything cannot be intelligent, it is just the hardware
 - The Robot Response:
 - a computer is void of senses and therefore symbols are meaningless to it, but a robot with sensors can tie its symbols to its senses and thus understand symbols
 - The Brain Simulator Response:
 - if we program a computer to mimic the brain (e.g., with a neural network) then the computer will have the same ability to understand as a human brain

12

Sci-Fi AI?


A collection of five images related to sci-fi AI: C-3PO and R2-D2; a Terminator robot; a woman's face next to a robot face; a man in sunglasses; and a small white robot head.

13

What is Artificial Intelligence ?

THOUGHT	Systems that think like humans	Systems that think rationally
BEHAVIOUR	Systems that act like humans	Systems that act rationally
	HUMAN	RATIONAL


15



Systems that **act** like humans

- For Turing, the cognitive tasks include:
 - *Natural language processing*
 - for communication with human
 - *Knowledge representation*
 - to store information effectively & efficiently
 - *Automated reasoning*
 - to retrieve & answer questions using the stored information
 - *Machine learning*
 - to adapt to new circumstances
- Ideally it includes two more issues, currently:
 - *Computer vision*
 - to perceive objects (seeing)
 - *Robotics*
 - to move objects (acting)


16



What is Artificial Intelligence ?

THOUGHT	Systems that think like humans	Systems that think rationally
BEHAVIOUR	Systems that act like humans	Systems that act rationally
	HUMAN	RATIONAL


17



Systems that **think** like humans

- Cognitive modeling
 - Humans as observed from 'inside'
 - Cognitive Science
 - Introspection vs. psychological experiments
 - "The exciting new effort to make computers think ... machines with *minds* in the full and literal sense" (Haugeland)
 - "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..." (Bellman)


18



What is Artificial Intelligence ?

THOUGHT	Systems that think like humans	Systems that think rationally
BEHAVIOUR	Systems that act like humans	Systems that act rationally
	HUMAN	RATIONAL


19



Systems that think 'rationally'

- "laws of thought"
 - Rational - defined in terms of logic?
 - Logic can't express everything (e.g. uncertainty)
 - Logical approach is often not feasible in terms of computation time (needs 'guidance')
- "The study of mental facilities through the use of computational models" (Charniak and McDermott)
- "The study of the computations that make it possible to perceive, reason, and act" (Winston)


20



What is Artificial Intelligence ?

THOUGHT	Systems that think like humans	Systems that think rationally
BEHAVIOUR	Systems that act like humans	Systems that act rationally
	HUMAN	RATIONAL

21



Systems that act rationally


- AI as a rational agent
 - It is more general than using logic only
 - LOGIC + Domain knowledge
 - Logic → only *part* of a rational agent, not *all* of rationality
 - Sometimes logic cannot reason a correct conclusion
 - At that time, some *specific (in domain) human knowledge* or information is used
 - It allows extension of the approach with more scientific methodologies
 - **Rational** behavior: doing the right thing
 - **The right thing**: that which is expected to maximize goal achievement, given the available information

22



TECHNICAL ASPECTS OF AI


29



So What Does AI Do?

- Most AI has fallen into one of two categories
 1. Select a specific problem to solve
 - study the problem (perhaps how humans solve it)
 - come up with the proper representation for any knowledge needed to solve the problem
 - acquire and codify that knowledge
 - build a problem solving system
 2. Select a category of problem or cognitive activity (e.g., learning, natural language understanding)
 - theorize a way to solve the given problem
 - build systems based on the model behind your theory as experiments
 - modify as needed
- Both approaches require
 - one or more representational forms for the knowledge
 - some way to select proper knowledge, that is, search

30

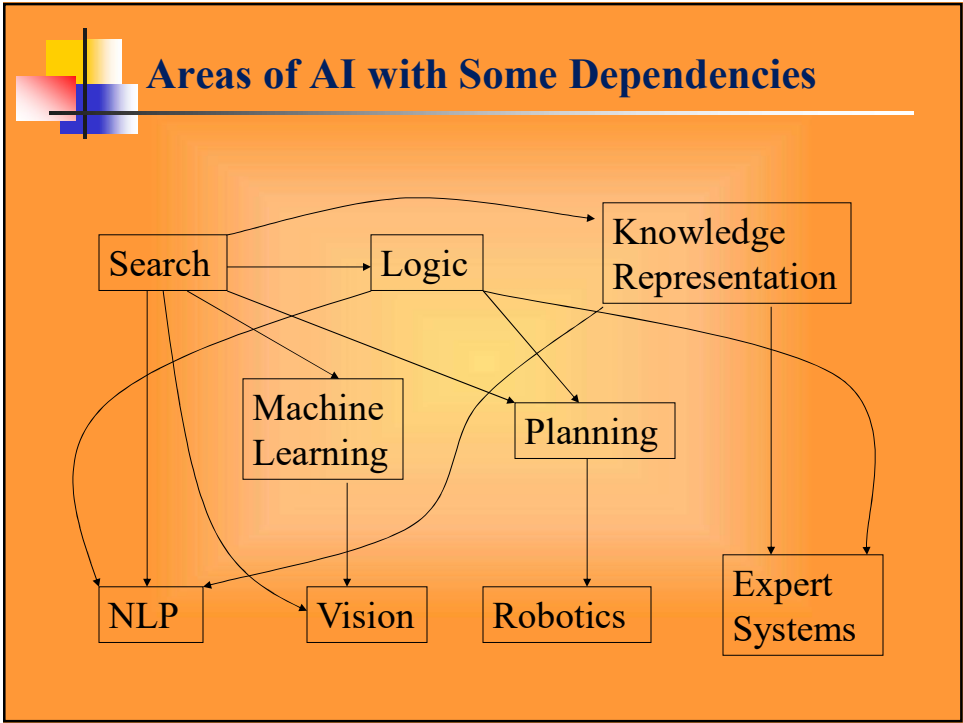


The main topics inside AI

Artificial intelligence can be considered under a number of headings:

- Search (includes Game Playing).
- Representing Knowledge and Reasoning with it.
- Planning.
- Learning.
- Natural language processing.
- Expert Systems.
- (now) Interacting with the Environment
(e.g. Vision, Speech recognition, Robotics)

31



32

Recall: what is Search?

- The state of the problem being solved = the values of the active variables
 - this will include any partial solutions, previous conclusions, user answers to questions, etc
- while humans are often able to make intuitive leaps, or recall solutions with little thought, the computer must search through various combinations to find a solution
- To the right is a search space for a tic-tac-toe game

33

Mixing concepts...

McCarthy: „[...] every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.“

ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt

MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time

DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data

Differentiation of AI, ML und DL
(picture source: Singh, Cousins of AI <<https://towardsdatascience.com/cousins-of-artificial-intelligence-dda4edc27b55>>)

34

ML versus DL

“Traditional” machine learning:

```
graph LR; CatImage[Cat Image] --> Handcrafted[handcrafted features]; Handcrafted --> Classifier[learned classifier]; Classifier --> CatOutput[cat]
```

Deep, “end-to-end” learning:

```
graph LR; CatImage[Cat Image] --> LowLevel[learned low-level features]; LowLevel --> MidLevel[learned mid-level features]; MidLevel --> HighLevel[learned high-level features]; HighLevel --> Classifier[learned classifier]; Classifier --> CatOutput[cat]
```

35



Relevance of Data

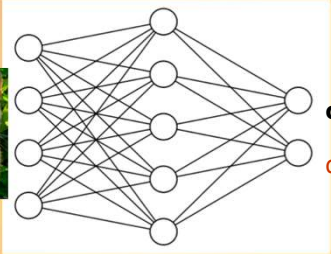

- Humans learn by observation and unsupervised learning
 - model of the world / common sense reasoning
- Machine learning needs lots of (labeled) data to compensate



36

Main types of machine learning

- **Supervised learning**
- Unsupervised learning
- Reinforcement learning



cat
dog

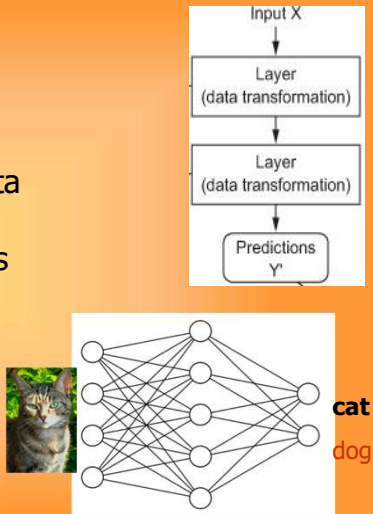
Two phases in the ML process:

- Training
- Evaluation/execution

37

Input data and targets

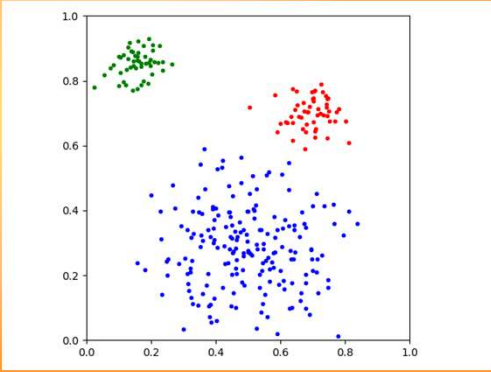
- The network maps the input data X to predictions Y'
- During **training**, the predictions Y' are compared to true targets Y using the loss function
- During **Evaluation**, the predictions are the outcomes of the system.



38

Main types of machine learning

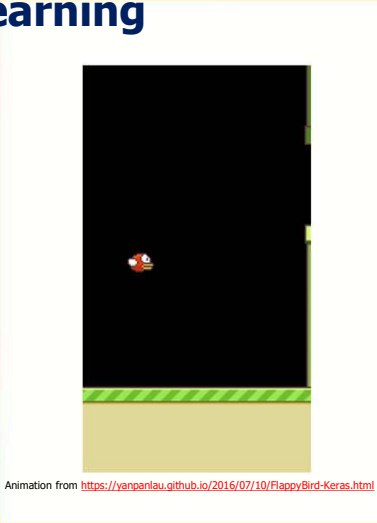
- Supervised learning
- **Unsupervised learning**
- Reinforcement learning



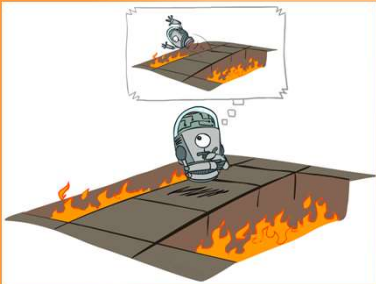
39

Main types of machine learning

- Supervised learning
- Unsupervised learning
- **Reinforcement learning**



Offline (MDPs) vs. Online (RL)




Offline Solution

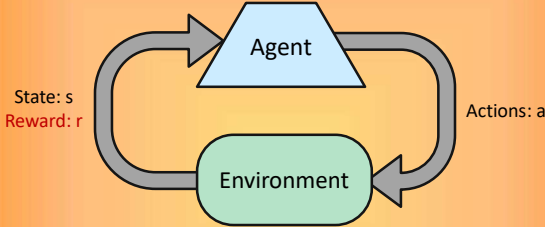


Online Learning

Offline – the training is not done with the system stopped
Online – the training is done while the system is operating




Reinforcement Learning



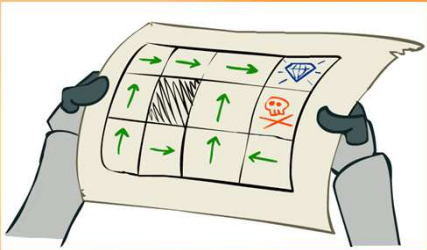
- Basic idea:
 - Receive feedback in the form of **rewards**
 - Agent's utility is defined by the reward function
 - Must (learn to) act so as to **maximize expected rewards**
 - All learning is based on observed samples of outcomes!

42



Passive Reinforcement Learning

- Simplified task: policy evaluation
 - Input: a fixed policy $\pi(s)$
 - You don't know the transitions $T(s,a,s')$
 - You don't know the rewards $R(s,a,s')$
 - Goal: **learn the state values**
- In this case:
 - Learner is "along for the ride"
 - No choice about what actions to take
 - Just execute the policy and learn from experience
 - This is NOT offline planning! You actually take actions in the world.



43

Active Reinforcement Learning

- Full reinforcement learning: optimal policies (like value iteration)
 - You don't know the transitions $T(s,a,s')$
 - You don't know the rewards $R(s,a,s')$
 - You choose the actions now
 - Goal: learn the optimal policy / values
- In this case:
 - Learner makes choices!
 - Fundamental tradeoff: exploration vs. exploitation
 - This is NOT offline planning! You actually take actions in the world and find out what happens...

A small, round, grey robot with a single eye and a small antenna is standing on a grey grid. The grid is surrounded by orange and yellow flames, indicating a hazardous environment. The robot appears to be exploring or navigating through the grid.


44

Direct Evaluation

- Goal: Compute values for each state under policy
- Idea: Average together observed sample values
 - Act according to policy
 - Every time you visit a state, write down what the sum of discounted rewards turned out to be
 - Average those samples
- This is called direct evaluation

A red slot machine with a digital display showing "DOUBLE \$2/\$0 OR NOTHING". The machine has two reels with symbols like cherries and diamonds. It has a coin slot on the right and a lever on the left.

45




Two AI Assumptions

1. We can *understand and model* cognition without understanding the underlying mechanism
 - It is the model of cognition that is important not the physical mechanism that implements it
 - If this is true, then we should be able to create cognition (mind) out of a computer or a brain or even other entities that can compute such as a mechanical device
 - This is the assumption made by symbolic AI
2. Cognition will emerge from the proper mechanism
 - The right device, fed with the right inputs, can learn and perform the problem solving that we, as observers, call intelligence
 - Cognition will arise as the result (or side effect) of the hardware
 - This is the assumption made by connectionist AI


While the two assumptions differ, neither is necessarily mutually exclusive and both support the idea **that cognition is computational**

46



AI – IS IT USED?

47




AI Applications

Other application areas:

- Gaming
- Robotics
- Bioinformatics:
 - Gene expression data analysis
 - Prediction of protein structure
- Text classification, document sorting:
 - Web pages, e-mails
 - Articles in the news
- Video, image classification
- Music composition, picture drawing
- Natural Language Processing 🗣️
- Perception.

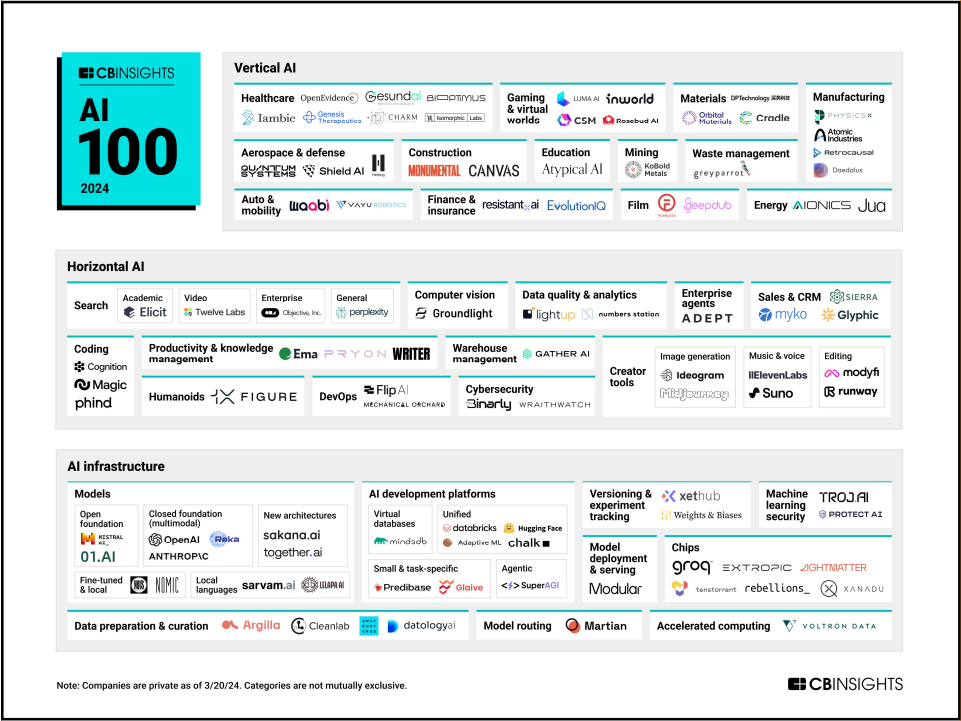
52




STATE OF AI/ML (2020)

Industry Wide	Higher Education
91.5% of leading businesses invest in AI	AI in the Education Market size exceeded \$1 billion in 2020.
97 million specialists needed in the AI industry by 2025	Expected to grow by more than 40% between 2021 and 2027.
Global AI market value is expected to reach \$267 billion by 2027.	As of 2020 only a minority of universities had an established AI strategy.
15.7 trillion influx into the global economy by 2030.	Most prevalent use of AI in Higher Ed is associated with academic integrity (plagiarism, proctoring).
Elimination of 85 million jobs and create 97 million new ones by 2025.	36% of Educause surveyed institutions use DAs and Chatbots. An additional 17% were either in planning or early implementation.
37% of businesses and organizations employ AI	22% of Educause surveyed institutions leverage AI for student success (academically at risk, early warnings).
2018 Gartner report predicted that through 2030, 85% of AI projects will provide false results caused by bias	Limited interest in leveraging AI for institutional activities such as curriculum planning.

53






What are AI problems *(class built)*

- Today?
 - Not 100% reliable
 - Associated costs (time, Money, CPU)
 - Loss of jobs
 - Deep fakes
 - Author/IPR rights
- Tomorrow?
 - Will reach the point in which is really indistinguishable from a reliable human
 - Create dependency on these systems
 - Even more loss of jobs
 - "cyborg" deep fakes

56

56



Learning

- If a system is going to act truly appropriately, then it must be able to change its actions in the light of experience:
 - how do we generate(?) new facts from old ?
 - how do we generate new concepts ?
 - how do we learn to distinguish different situations in new environments ?
 - How do we learn while we are acting?

57

Users, Data and Algorithms

With online learning, there is a clear danger that bias become self-fullfilling

60

60

Example: Bias on images

Fig. 4. Geographic distribution of countries in the Open Images data set. In their sample, almost one third of the data was US-based, and 60% of the data was from the six most represented countries across North America and Europe, from [142] © Shreya Shankar.


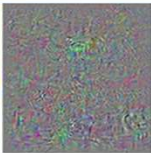


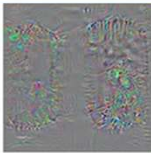

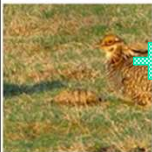
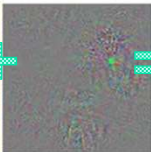
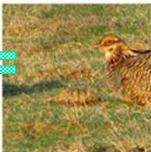

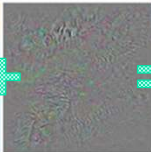


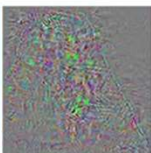


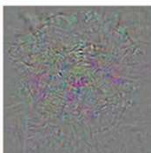

How well trained will be images from other countries?
How will that affect the classification/action of the AI?

61

61

Vulnerabilities on evaluation





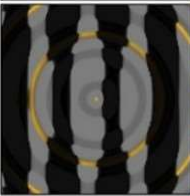

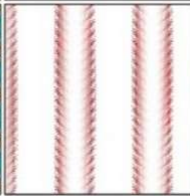
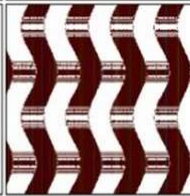
"adding invisible" distortion on the image deeply changes the classification

					
					
					
correct	+distort	ostrich	correct	+distort	ostrich

<http://karpathy.github.io/2015/03/30/breaking-convnets/>

62

Vulnerabilities on evaluation: complete mismatch


			
robin	cheetah	armadillo	lesser panda
			
king penguin	starfish	baseball	electric guitar

Note: this problem can be useful/problematic/intended:


- i) embedded during training,
- ii) failure during evaluation or
- iii) added for privacy during evaluation

63

63



(Mis)evaluation impact in real life




Think:

- Employment impact
- Insurance impact
- Reputation impact

(picture source: Elliott, AI Cartoons
<<https://timoelliott.com/blog/cartoons/artificial-intelligence-cartoons>>)

64



AI and Security

- Attack AI systems
 - Learning
 - Cause learning system to not produce intended/correct results
 - Cause learning system to produce targeted outcome designed by attacker
 - System/learning
 - Learn sensitive information about individuals

⇒ Need security in learning systems
- Misuse AI
 - Use AI to attack other systems
 - Find vulnerabilities in other systems
 - Target attacks
 - Devise attacks

⇒ Need security in other systems

65



How to overcome these AI social issues

Who can regulate the use of AI?

- **European law:** if there is a reference to the internal market and thus a need for legal harmonisation: e.g. differences between national AI regulations make cross-border activities more burdensome
- **International law** (e.g. "European Ethical Charter on the use of AI in judicial systems and their environment" of the Council of Europe)
- **National law**
- **Professional codes** - self-regulation as a "privilege" of the liberal professions

68



High-Level Expert Group on AI: Ethics Guidelines

Created to drive “responsible” AI usage

4 ethical principles:	7 core requirements:
<ol style="list-style-type: none">1. Respect for human autonomy2. Prevention of harm3. Fairness4. Explicability	<ol style="list-style-type: none">1. Human agency and oversight2. Technical robustness and safety3. Privacy and Data Governance4. Transparency5. Diversity, non-discrimination and fairness6. Societal and environmental wellbeing7. Accountability

(source: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines>)

69



AI and human rights: Charter of Fundamental Rights, ECHR, constitutions

- **Responsibility for the consequences of innovation:**
 - The state guarantees protection from negative effects of technological innovation
 - Principle of non-discrimination – **Attention: correlation instead of causality**
- **Freedom of innovation:** Securing the freedom for technical development - Freedom to conduct business, right to (intellectual) property

(Example: Necessary standard of medical treatments: Obligation to use AI?
(e.g. ECHR 30.8.2016, 40448/06 *Aydoğdu/Turkey*: functioning hospital system)