



## Last lessons....

- AI
  - Market, technology, what is “IA”
- Open source models, trabalhos derivados
- Marcas, IPR
- Standards
- Dilemas de personalidade – tempo e constância.

1

1



## Aspetos Profissionais e Sociais da Engenharia Informática

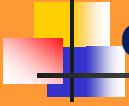
**This is not a philosophy lesson...**

*And practical works*

Rui L Aguiar, UA/IT

2

1



## Objective of this class

- Notions on several concepts
  - Applied AI issues: autonomous driving and decisions
  - Review Work #1
  - Discussion Work #2

3

3



## Law, Regulations, Ethics

- You cannot avoid following the law.
  - There are consequences
- Regulations are professional aspects that must/should/may be followed in a specific context
  - Enforced by different (professional) bodies
- You can avoid following ethics, but may be community consequences
  - Anyone remembers “cancelling”? “untrustable partner”?

4

4

Lets have fun - II

## HUMAN AGENCY

5

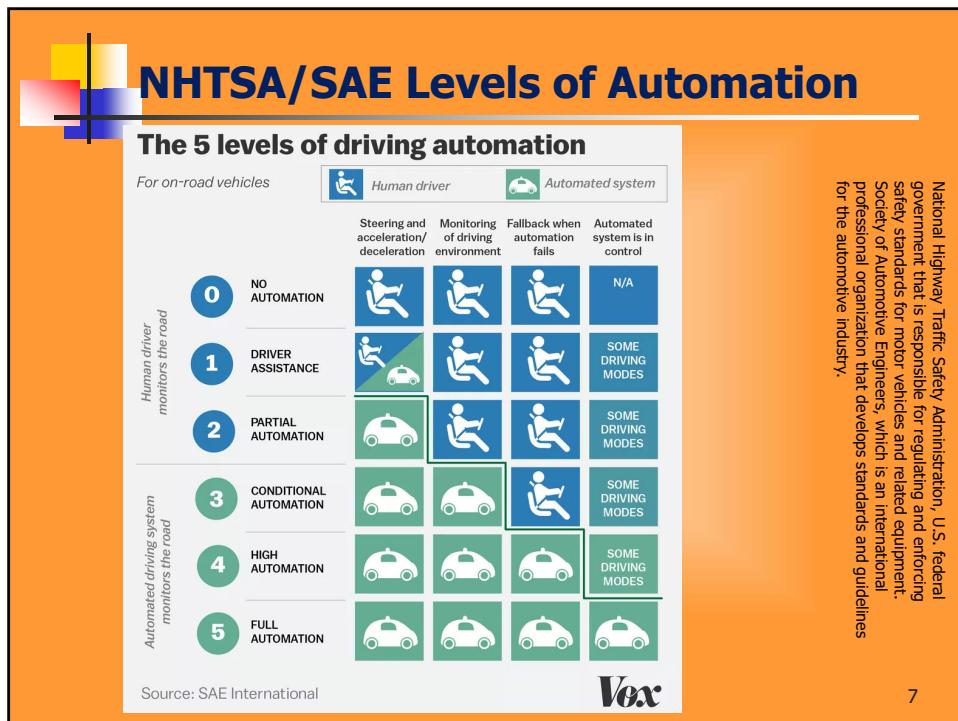
### Context – why not Automated Driving?

#### The five stages of autonomy

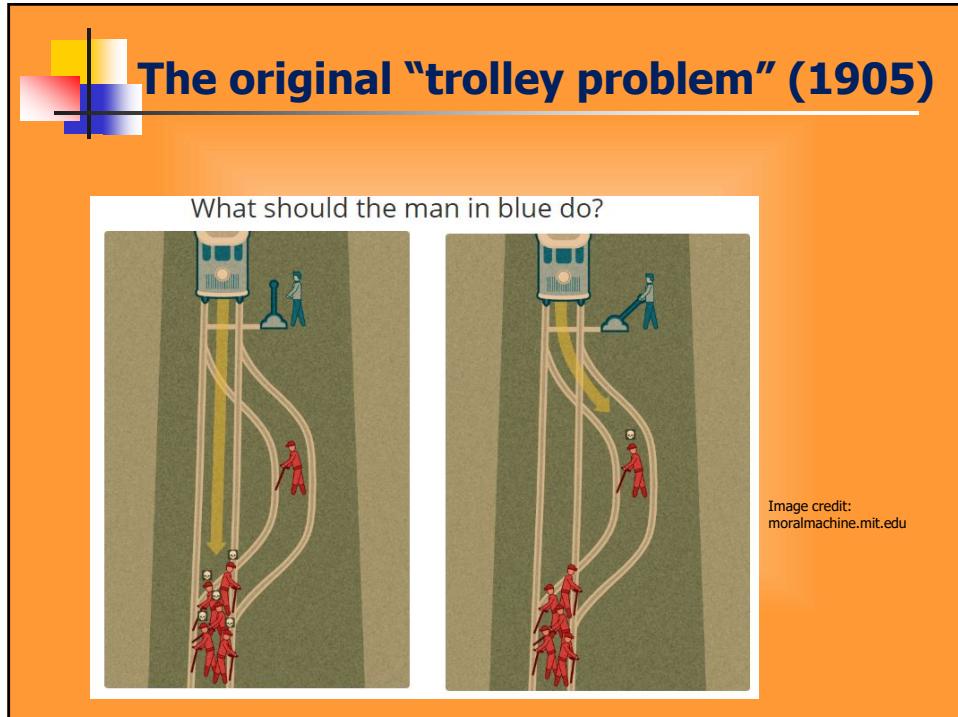


6

6



7



8



9

<ul style="list-style-type: none"><li>■ Move lever<ul style="list-style-type: none"><li>■ Menos pessoas morrem</li><li>■ Ser considerado herói</li><li>■ Em 5 pessoas tem mais probabilidades de "não ser mal" do que de uma pessoa</li><li>■ Se eu souber que pessoas estão em cada linha afeta a minha decisão</li></ul></li></ul>	<ul style="list-style-type: none"><li>■ Do not move lever<ul style="list-style-type: none"><li>■ Não assumir decisão</li><li>■ Não me envolver no problema em questão</li><li>■ Sentido de responsabilidade: não devia estar na linha a dormir</li><li>■ Mudar implica saber que é preferível</li><li>■ Se puxasse a alavanca teria de ser consistente no resto da vida</li><li>■ Se eu souber que pessoas estão em cada linha afeta a minha decisão (podia matar a pior pessoa do mundo ou a melhor)</li></ul></li></ul>
--	---

10

10



## Reasons – class built in 2024!

- Move lever
  - Principle of maximum happiness
  - Minor certainty on the outcome
  - Making people responsible per actions (sleeping person)
- Do not move lever
  - Unable to take decision
  - Not accepting responsibility
  - Fate

11

11

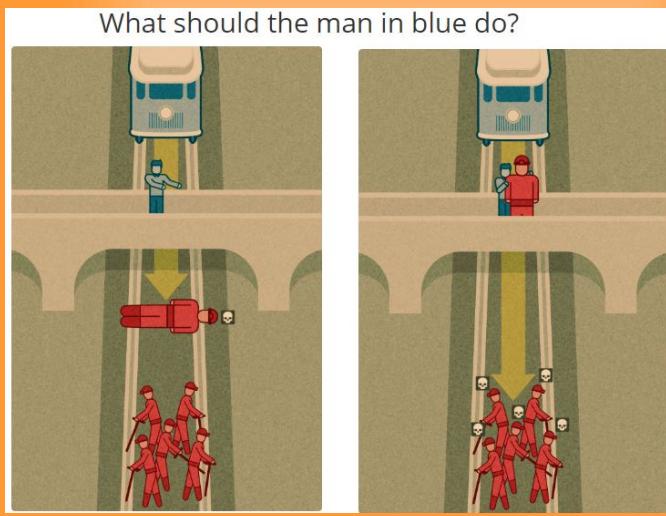


## Reasons – class built in 2023!

- Move lever
  - It is the smallest evil
  - It is the rationale for utility function (a-la Stuart Mill)
  - Assuming:
    - All persons are equal
    - I do not know anyone in the scenario
- Do not move lever
  - It is a crime
  - I cannot make judgments on people I do not know
  - It is not my right to decide who lives and who dies

12

12



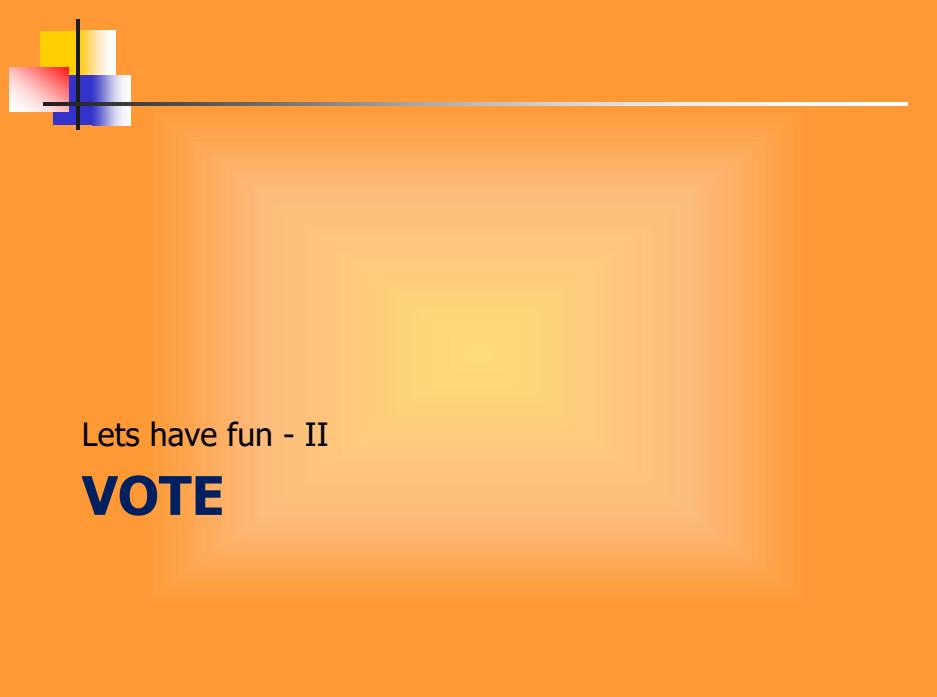
The original “trolley problem” (1905)

What should the man in blue do?

A diagram illustrating the trolley problem. On the left, a blue trolley is heading towards five people tied to the tracks. A man in a blue shirt stands near a switch or lever. On the right, a red trolley is heading towards five people tied to the tracks. A man in a red shirt stands near a switch or lever. In both cases, if the man pushes the lever, the trolley will divert to a track with only one person tied to it.

Image credit:  
moralmachine.mit.edu

13



Lets have fun - II

**VOTE**

14



## Reasons – class built!

■ Push person

- Menos pessoas morrem
- Ser considerado herói
- Em 5 pessoas tem mais probabilidades de “não ser mal” do que de uma pessoa
- É igual ao problema anterior, mas com mais na consciência.

■ Do not push person

- A pessoa que está na ponte não se colocou em perigo, e eu não tenho o direito de tomar a decisão de a sacrificar
- Podia saltar eu
- Humanismo
- Não assumir decisão
- Não me envolver no problema em questão
- Sentido de responsabilidade: não devia estar na linha a dormir
- Mudar implica saber que é preferível
- Se puxasse a alavanca teria de ser consistente no resto da vida

15

15



## Reasons – class built 2024!

■ Push person

- Principle of maximum happiness
- Minor certainty on the outcome
- Making people responsible per actions (sleeping person)

■ Do not push person

- Unable to take decision
- Not accepting responsibility
- Fate
- Making people responsible
- Personal perception associated in action
- The tool (the person) is innocent.

16

16



## Reasons – class built in 2023!

- Push person

- It is the smallest evil
- It is the rationale for utility function (a-la Stuart Mill)
- Assuming:
  - All persons are equal
  - I do not know anyone in the scenario

- Do not push person

- It is a even larger crime
- I cannot make judgments on people I do not know
  - It is not my right to decide who lives and who dies
  - The person on the bridge has no fault on the situation

17

17



## Had fun?

- Have you heard of police and the courts?
  - In general...
  - You will be held strongly accountable of ACTING
  - You will be less chargeable (if at all) for FAILING to ACT.

Details will be important: are you driving the trolley? Are you watching? Were you able to think it over? Are you related with any person in question?

Side note: there are different philosophy theories associated to these cases (*utilitarianism, intention, etc..*)

18

## MIT's Moral Machine

- MIT moral experiment for automotive drive
- Multiple scenarios, multiple personas
- 2 million people, 10 languages, 233 countries = 40 million decisions
- In the main interface of the Moral Machine, users are shown unavoidable accident scenarios with two possible outcomes, depending on whether the autonomous vehicle swerves or stays on course
- **Results were VERY VARIANT of country and culture!**

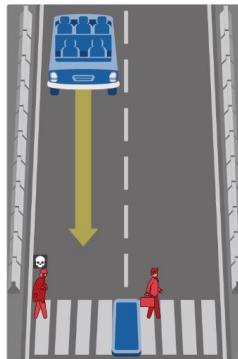
19

## The Moral Machine

What should the self-driving car do?

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in...

- Dead:
- 1 homeless person



2 / 13

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in...

- Dead:
- 1 male executive

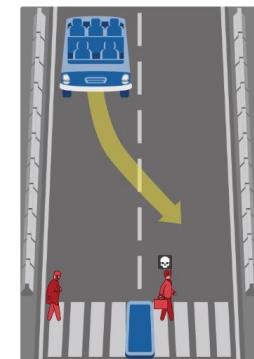


Image credit:  
mormachine.mit.edu

20

## The Moral Machine

What should the self-driving car do?

In this case, the self-driving car with sudden brake failure will continue ahead and crash into a concrete barrier. This will result in ...

Dead:

- 1 female athlete
- 2 male athletes

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...

Dead:

- 1 large woman
- 2 large men

Image credit: [moralmachine.mit.edu](http://moralmachine.mit.edu)

21

## The Moral Machine

What should the self-driving car do?

In this case, the self-driving car with sudden brake failure will continue ahead and crash into a concrete barrier. This will result in ...

Dead:

- 1 baby
- 1 large woman

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...

Dead:

- 1 baby

Note that the affected pedestrians are abiding by the law by crossing on the green signal.

Image credit: [moralmachine.mit.edu](http://moralmachine.mit.edu)

22

## The Moral Machine

What should the self-driving car do?

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...

**Dead:**

- 3 girls
- 1 female doctor
- 1 female executive

Note that the affected pedestrians are flouting the law by crossing on the red signal.

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing in the other lane. This will result in ...

**Dead:**

- 3 boys
- 1 male doctor
- 1 male executive

Note that the affected pedestrians are abiding by the law by crossing on the green signal.

Image credit: [moralmachine.mit.edu](http://moralmachine.mit.edu)

Copyright Tara Goddard  
2019

23

## The Moral Machine

What should the self-driving car do?

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...

**Dead:**

- 1 woman
- 1 large woman
- 1 female executive
- 1 large man
- 1 girl

Note that the affected pedestrians are abiding by the law by crossing on the green signal.

In this case, the self-driving car with sudden brake failure will continue ahead and crash into a concrete barrier. This will result in ...

**Dead:**

- 1 elderly woman

Image credit: [moralmachine.mit.edu](http://moralmachine.mit.edu)

24

## The Moral Machine

What should the self-driving car do?

In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in ...

**Dead:**

- 1 male athlete
- 1 male doctor

Note that the affected pedestrians are abiding by the law by crossing on the green signal.

8 / 13

In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...

**Dead:**

- 1 cat
- 1 dog

Note that the affected pedestrians are flouting the law by crossing on the red signal.

Image credit: [moralmachine.mit.edu](http://moralmachine.mit.edu)

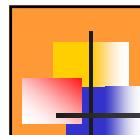
25

## Lets return to automotive driving

<< Please don't make me choose!  
Just tell me what to do. >>

~10/11  
2-2021

26



## Easy for an engineer...

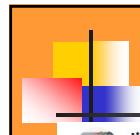
- Engineers are better than these abstract dilemmas, right?
  - Do you plan for “brake failures” which may lead to “unavoidable accidents” ?
  - Do you drive fast enough in the vicinity of pedestrian crossings that they would kill people in the case of a crash? And if you are in a rush?

(Have you heard of “Planned obsolescence”?)

*“The only safe scenario would be don’t move. You have to make reasonable assumptions about what you care about and what you don’t.”*

*Aaron Ames.*

27



## Do you remember that “law” thingy?

Jim McPherson @SafeSelfDrive Following

Contract.

**Navigate on Autopilot (Beta)**

When Navigate on Autopilot is enabled your Model 3 will determine which lane you need to be in and when. In addition to ensuring you reach your intended exit, Autopilot will watch for opportunities to move to a faster lane when you’re caught behind slower traffic. When you reach your exit, your Model 3 will depart this freeway, slow down and transition control back to you.

Use Navigate on Autopilot only if you will pay attention to the road, keep your hands on the steering wheel, and be prepared to take over at any time. Your Model 3 will indicate when a lane change or exit is coming, but you are still responsible for monitoring the environment and maintaining control.

When exiting highways, remember that Autopilot will not stop on its own at stop lights or stop signs, or yield for merges.

Do you want to enable Navigate on Autopilot while it is in Beta?

NO YES

If-Driving Capability

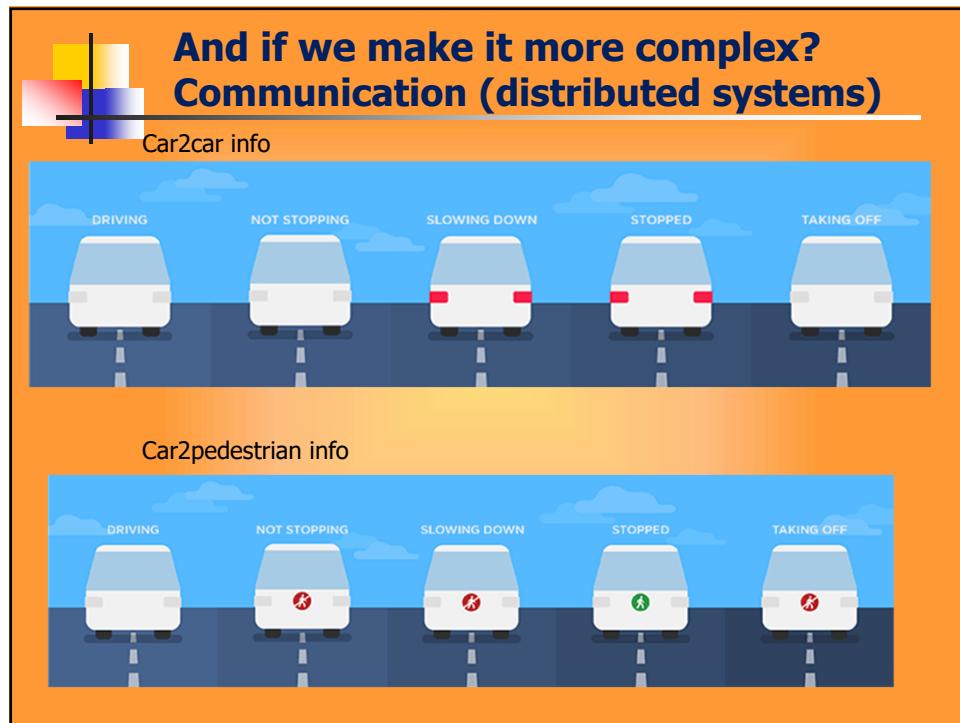
that enhances safety and convenience behind the overall workload as a driver. Each new Tesla vehicle will feature enhanced vision processing to provide an additional layer of safety and reliability. Middle Eastern markets will now utilise our cameras with radar and instead rely on Tesla's advanced suite of autopilot and related features.

Autopilot comes standard on every new Tesla. Owners who took delivery of their cars without Autopilot, there are two Autopilot packages available for purchase, depending on when your car was built: Enhanced Autopilot and Full Self-Driving Capability.

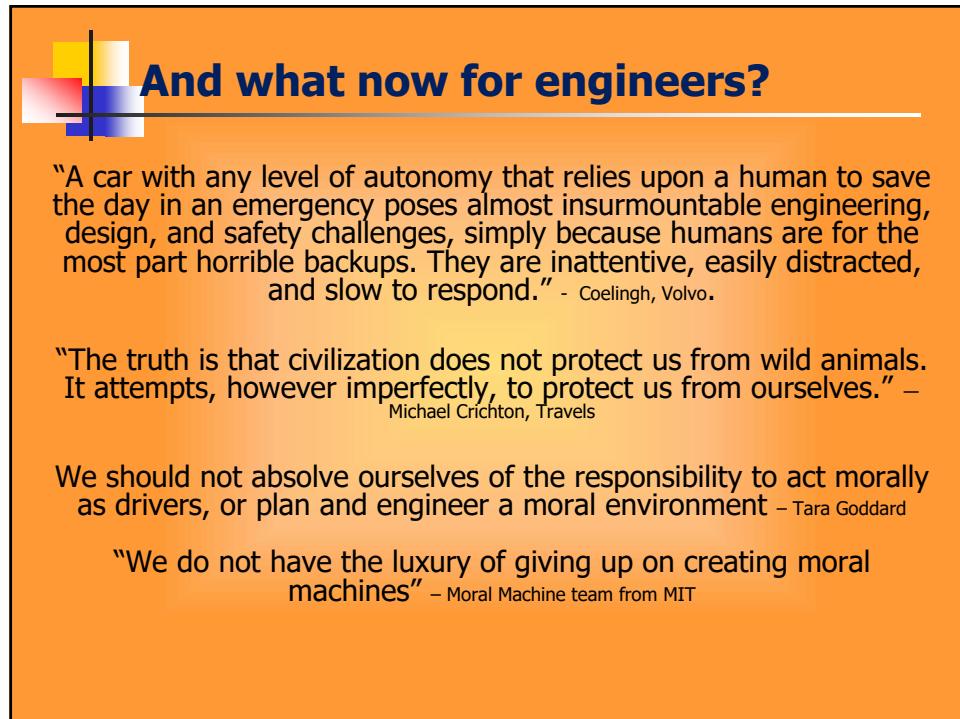
Autopilot, Enhanced Autopilot and Full Self-Driving Capability are intended for use with a fully attentive driver, who has their hands on the wheel and is prepared to take over at any moment. While these features are designed to become more capable over time, the currently enabled features do not make the vehicle autonomous.

28

28



29



30

## Legal – Informatics ?

Law ?

31

### (Mis)evaluation impact in real life



*"The good news is that you're perfectly healthy —the bad news is that my algorithms predict you'll be dead in two days..."*

Think:

- Employment impact
- Insurance impact
- Reputation impact

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

32



## 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

33



## 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

34



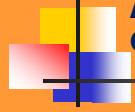
## High-Level Expert Group on AI: Ethics Guidelines

Created to drive “responsible” AI usage

4 ethical principles:	7 core requirements:
1. Respect for human autonomy 2. Prevention of harm 3. Fairness 4. Explicability	1. Human agency and oversight 2. Technical robustness and safety 3. Privacy and Data Governance 4. Transparency 5. Diversity, non-discrimination and fairness 6. Societal and environmental wellbeing 7. Accountability

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

35

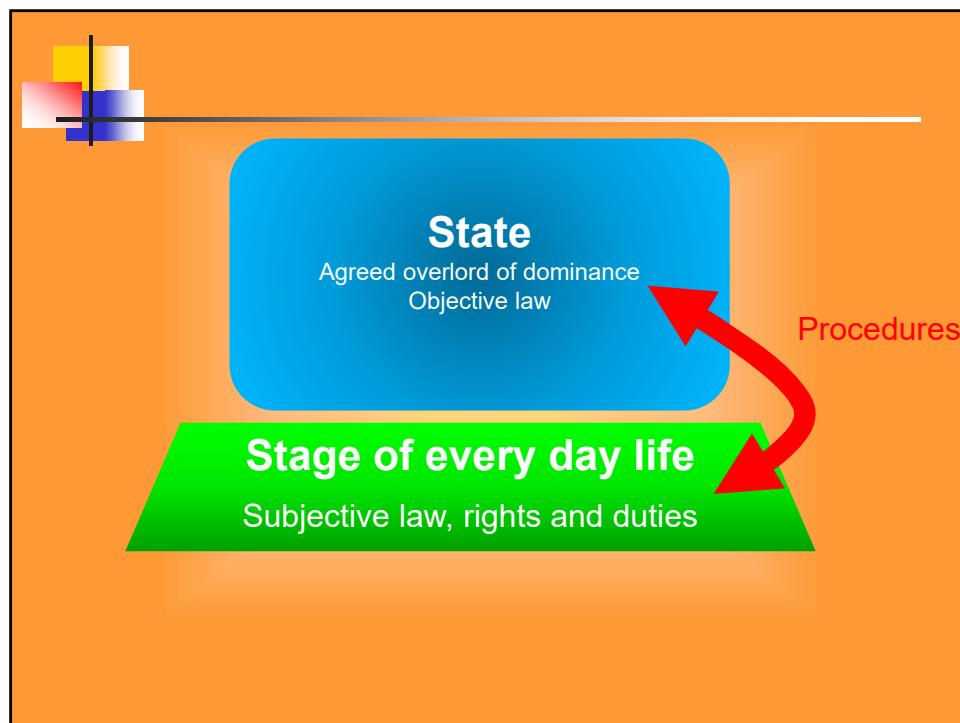


## AI and human rights: Charter of Fundamental Rights, ECHR, constitutions

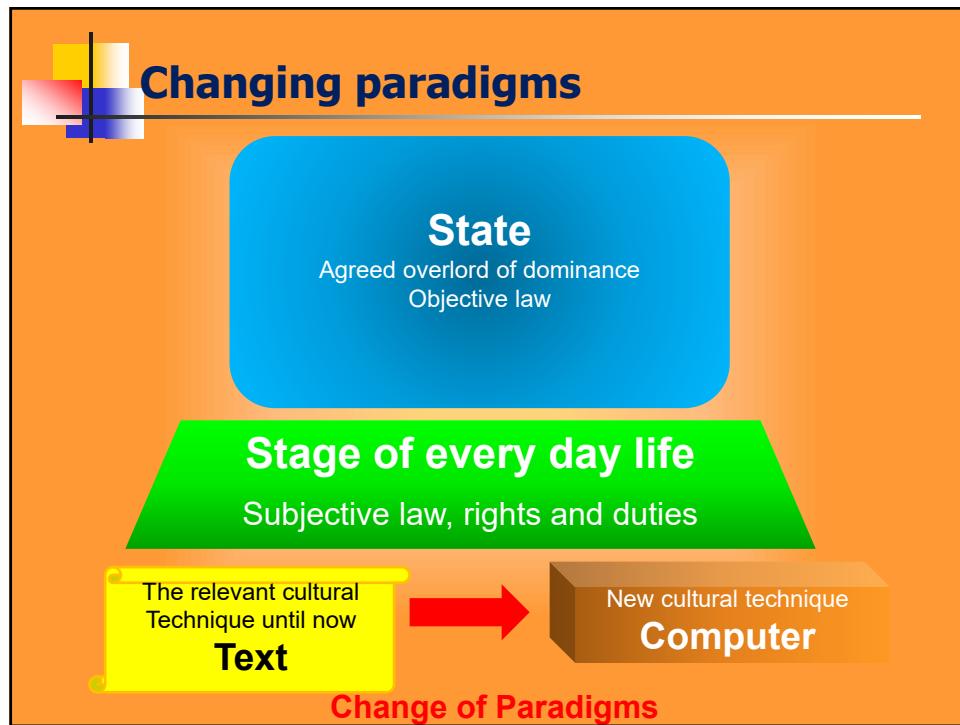
- **Responsibility for the consequences of innovation:**
  - The state guarantees protection from negative effects of technological innovation
  - Principle of non-discrimination – **Attention: correlation instead of causality**
- **Freedom of innovation:** Securing the freedom for technical development - 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)

36



37



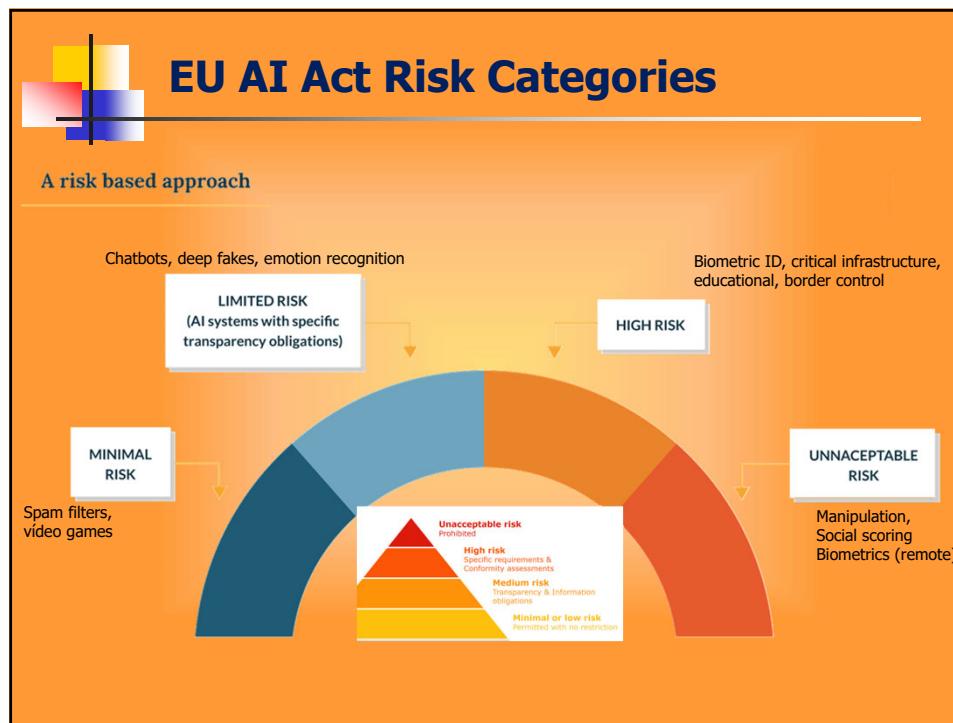
38

## EU AI Act

- Two main concepts
  - What is AI
  - Risk based-approach
- Enforced by EC and national authorities
  - Penalties may reach 6% turnaround worldwide
  - Potential AI Board and AI Office (Europe)
    - Code of conduct

39

39



40



41

## WHAT ABOUT THE FUTURE?

42

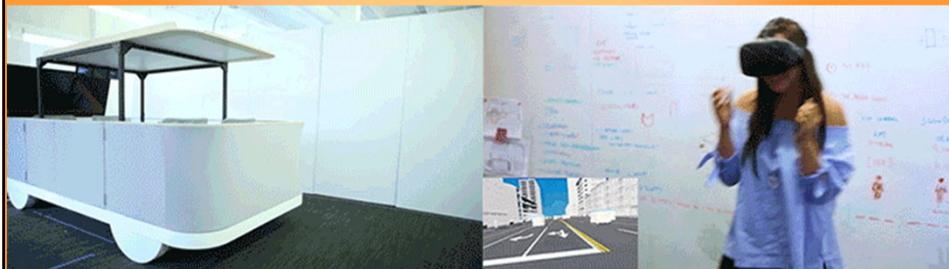
## Deep challenges in Informatics

- The definition of Identity
  - When is me really me?
  - I can clone myself digitally
- The association of a Digital Identity to a (real or virtual) object
  - Is this permanent? Can be transitive?
  - E.g. the DNS system works along these lines. The CC is another such system
- The structure and definition of the *Digital Identity identifier*

43

43

## Lets' return to the autonomous car Virtual reality



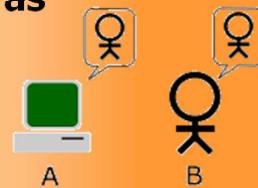
Do you understand why the concepts of identity and identifier are so important?  
This concept is known as *digital twin*.

Source: Crossing the road in the world of autonomous cars by Stamp Siripanich at Teague Labs <https://link.medium.com/rLZQ1gC6GV>

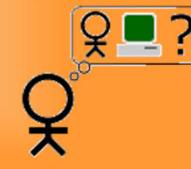
44

## Another example: “intelligence” a-la Turing Test

- You ask the computer any question.
- If it is able to reply to it in the same manner a normal human would, **it is named as “intelligent”.**



- Turing test: chat room.
  - Many participants in a chat room.
  - One of the participants is a real person and one participant is a computer program.
  - The program passes the test (is considered intelligent) if no one can tell which of the two participants is human.

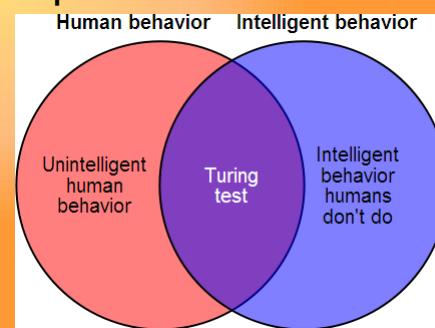


45

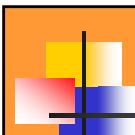
## But naming is not the identity

- Turing test does not directly test if the computer behaves intelligently
- It only tests whether or not the computer behaves like a human being.
- In general the two concepts are different.

Understanding this separation between the digital world and the real world, and how we relate both is a fundamental tenant in informatics



46



## Will AI Be Like Us?

**Like us, AI systems...**

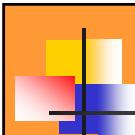
- ...will talk to us in our languages.
- ...will help us with our problems.
- ...will have anthropomorphic interfaces.

**Unlike us, AI systems...**

- ...will compute and communicate extremely quickly.
- ...will have bounds for learning and retention of knowledge that will soon surpass ours.
- ...might not be well modeled by the psychological models that work for people.

47

47



## Tools vs Agents

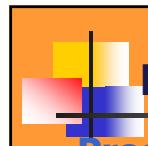
**Agent:** Takes responsibility, takes initiative, interacts with others on behalf of a client.

**Tool:** Responds directly to its user. Does not take responsibility. Does not take initiative. Does not normally interact with others on behalf of a client.

What will be the outcome?

48

48

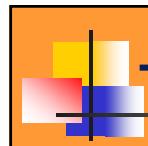


## High Levels of AI

- Pros...
  - Powerful tools,
  - solutions to complex problems,
  - better society(?)
  
- Cons...
  - As a tools it might be used against people;
  - May create worse problems than it solves;
  - standards of living might get worse;
  - *we might lose some aspect of our humanity.*

49

49



## The Safe Systems approach

- Autonomous and semi-autonomous technologies should be developed within a Safe Systems framework, NOT the other way around
- Ex: automotive vision zero approach:

<b>TRADITIONAL APPROACH</b> <p>Traffic deaths are <b>INEVITABLE</b>  <b>PERFECT</b> human behavior        Prevent <b>COLLISIONS</b>        INDIVIDUAL responsibility        Saving lives is <b>EXPENSIVE</b></p>		<b>VISION ZERO</b> <p>Traffic deaths are <b>PREDICTABLE</b>        Integrate <b>HUMAN FAILING</b> in approach        Prevent <b>FATAL AND SEVERE CRASHES</b>        SYSTEMS approach        Saving lives is <b>NOT EXPENSIVE</b></p>
--	---	--

Copyright Tara Goddard 2019

50



## Food for thought: intelligence and identity

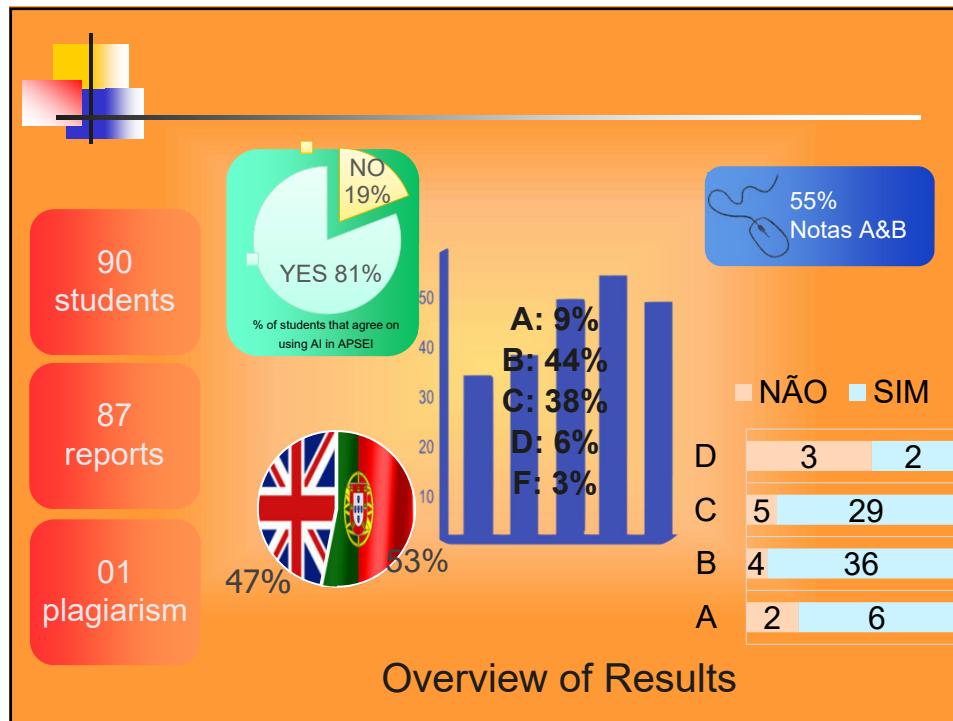
- Metaverse
  - Entities in the metaverse
  - Digital currencies
  - Non-fungible-tokens
- Services/Programs
  - All the communication stack!
  - All the process management!
- Legal frameworks rely in these concepts!
  - Uniqueness concepts
  - Digital world is challenging laws and regulations!

51



Work #1

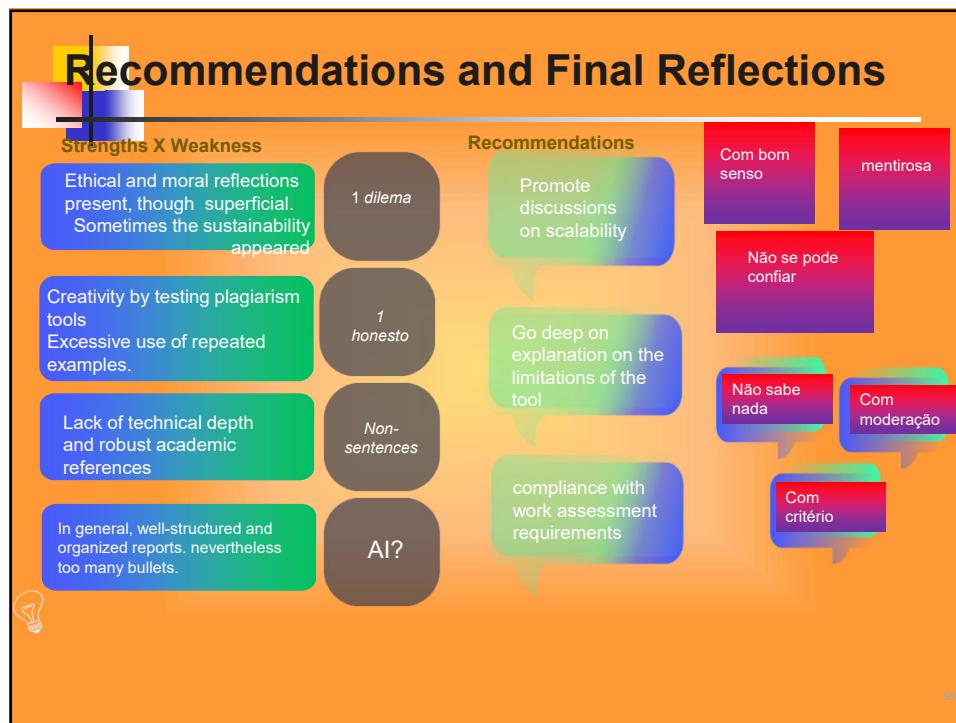
52



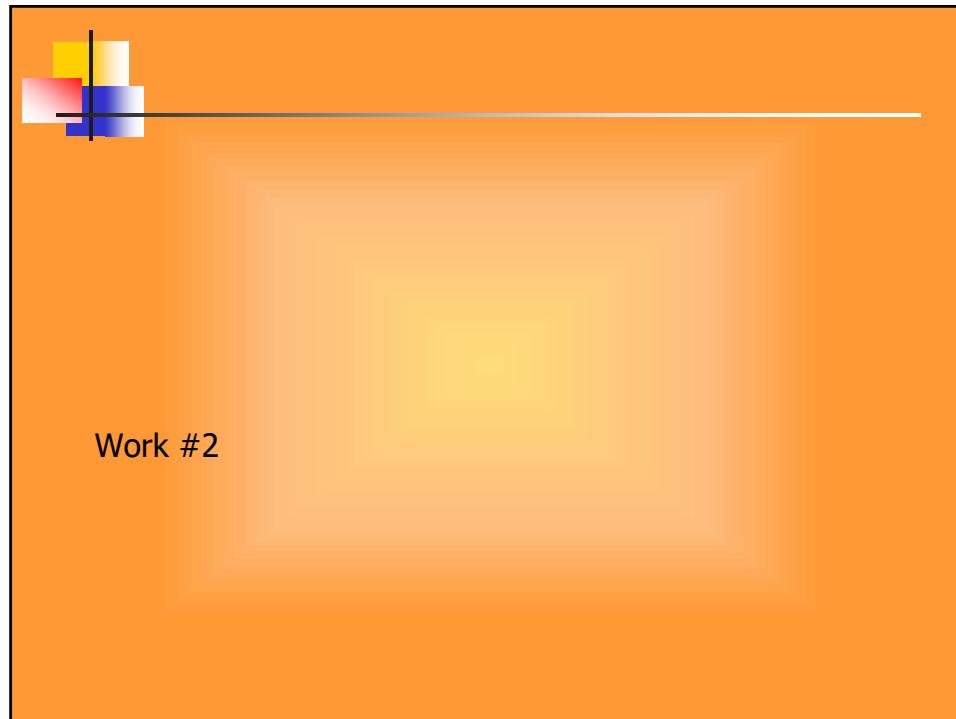
53



54



55



56



## Assignment 2 – Business Assessment

- 1. You will create your own Startup on the AI market. Choose your positioning, but it must be associated to a very large number of final users (not necessarily customers).**
- 2. Make a sketch of the needs for such a company, if it becomes very successful**
  - 1. Which departments you will need**
  - 2. To execute which functions,**
  - 3. With how many employees**
  - 4. With what cost estimates**
  - 5. What resources you will need**
- 3. Prepare a timeline for moving into market, from the initial idea, until MVP completed.**

***The work is to be done in groups of two students.***

***Deadline 27<sup>th</sup> March. Evaluation criteria presented in next class***

57

57



## Grading criteria

- **Delivery:**
  - 1 min pitch of the idea and company structure
  - Excel spreadsheet with costs and company timeline with objectives
  - Document (4 pages) explaining these costs.
- **Parameters:**
  - Completeness of the answers
    - Dimensions considered for the company: how realistic it is, and how the different dimensions of a large AI company in EUROPE are properly considered
    - Proper identification of the assumptions made on the answers, and how reasonable they seem.
  - Scalability
    - Discussion on the developments of the company, its scaling, and varying costs as company increases
  - Technological reasoning
    - Technical aspects that are relevant for supporting the needs/costs identified in the other points
    - Pre/pos supply chain technology aspects
  - Quality of delivery
    - Indication of sources, language proficiency (both PT and UK are accepted), overall professional delivery
- All points to be (soft) graded: D,C,B,A

58

58