




Aspetos Profissionais e Sociais da Engenharia Informática

This is not a philosophy lesson...
but will take some time for you to understand the title



Rui L Aguiar, UA/IT

2

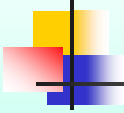


Objective of this class

- Notions on several concepts
 - Ethics
 - Identity
 - Impact in Informatics


3

3









Lets have fun - I

4



Context – why not Automated Driving?

The five stages of autonomy

0. DRIVER	1. FEET OFF	2. HANDS OFF	3. EYES OFF	4. MIND OFF	5. PASSENGER
					
No assistance	Assisted	Patially automated	Highly automated	Fully automated	Autonomous
Human	Transfer of responsibility				Machine

Sources: Evercore ISI, SAE International

5

5

NHTSA/SAE Levels of Automation

The 5 levels of driving automation

For on-road vehicles

Human driver

Automated system

	Steering and acceleration/deceleration	Monitoring of driving environment	Fallback when automation fails	Automated system is in control
<div>0</div> <div>NO AUTOMATION</div>				N/A
<div>1</div> <div>DRIVER ASSISTANCE</div>				SOME DRIVING MODES
<div>2</div> <div>PARTIAL AUTOMATION</div>				SOME DRIVING MODES
<div>3</div> <div>CONDITIONAL AUTOMATION</div>				SOME DRIVING MODES
<div>4</div> <div>HIGH AUTOMATION</div>				SOME DRIVING MODES
<div>5</div> <div>FULL AUTOMATION</div>				

Human driver monitors the road

Automated driving system monitors the road

Source: SAE International

Vox

National Highway Traffic Safety Administration, U.S. federal government that is responsible for regulating and enforcing safety standards for motor vehicles and related equipment. Society of Automotive Engineers, which is an international professional organization that develops standards and guidelines for the automotive industry.

6

6


The original “trolley problem” (1905)

What should the man in blue do?

Image credit: moralmachine.mit.edu

7


3



Lets have fun - I

VOTE

8



Reasons – class built!

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

- Do not move lever
 - Unable to take decision
 - Not accepting responsibility
 - Fate

9

9

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

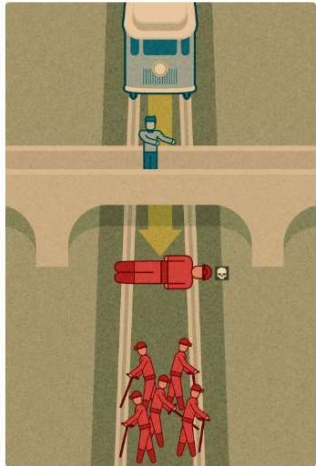
■ It is not my right to to decide who lives and who dies

10

10

The original “trolley problem” (1905)

What should the man in blue do?




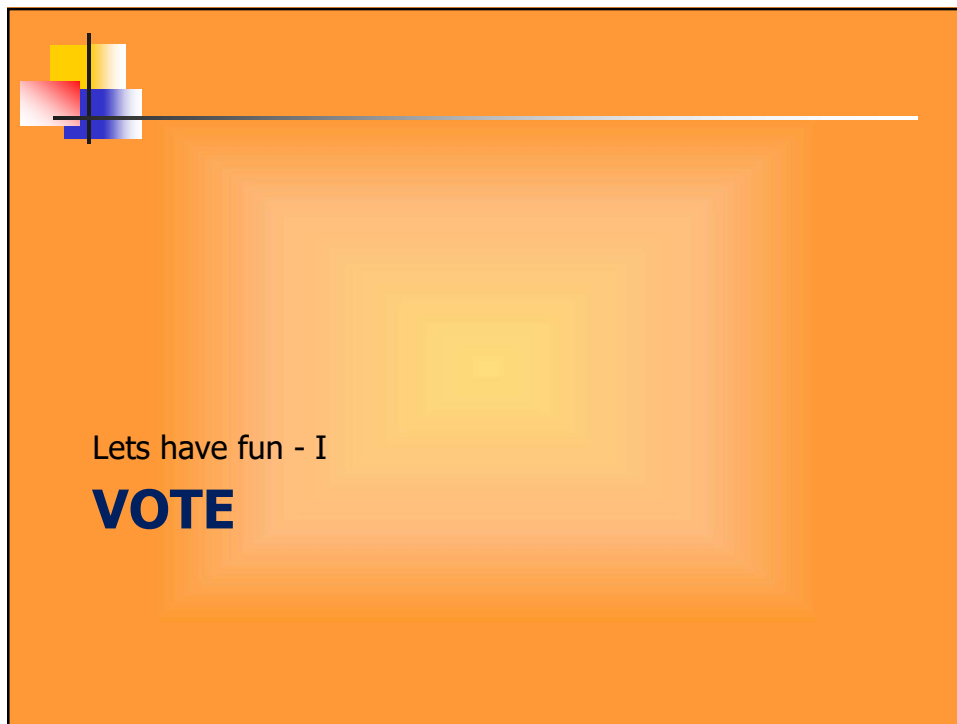


Image credit: moralmachine.mit.edu

11




12

Reasons – class built!

<ul style="list-style-type: none">■ Move lever<ul style="list-style-type: none">■ Principle of maximum hapiness■ Minor certainty on the outcome■ Making people responsible per actions (sleeping person)	<ul style="list-style-type: none">■ Do not move lever<ul style="list-style-type: none">■ Unable to take decision■ Not accepting responsibility■ Fate■ Making people responsible■ Personal perception associated in action■ The tool (the person) is innocent.
--	--

13

13




Reasons – class built in 2023!

<ul style="list-style-type: none"> ■ Push person <ul style="list-style-type: none"> ■ It is the smallest evil ■ It is the rationale for utility function (a-la Stuart Mill) ■ Assuming: <ul style="list-style-type: none"> ■ All persons are equal ■ I do not know anyone in the scenario 	<ul style="list-style-type: none"> ■ Do not push person <ul style="list-style-type: none"> ■ It is a even larger crime ■ I cannot make judgments on people I do not know <ul style="list-style-type: none"> ■ It is not my right to to decide who lives and who dies ■ The person on the bridge has no fault on the situation
---	--

14

14



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..*)

15

15



MIT's Moral Machine

- MIT moral experiment for automative 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!**



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


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

2 / 13

Image credit: moralmachine.mit.edu



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

4 / 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 large woman
- 2 large men

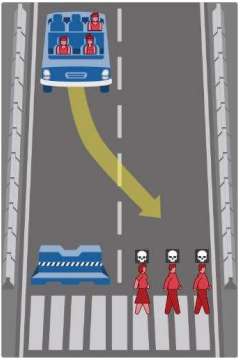
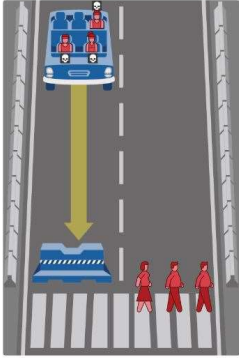



Image credit: moralmachine.mit.edu



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

1 / 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 baby

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

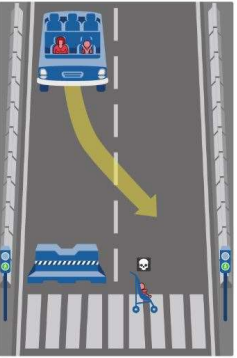
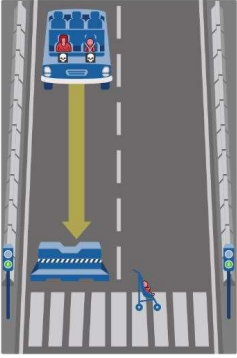



Image credit: moralmachine.mit.edu



The Moral Machine

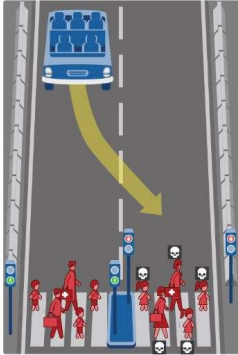
In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...

Dead:

- 3 girls
- 1 female doctor
- 1 female executive

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

What should the self-driving car do?



3 / 13

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:

- 3 boys
- 1 male doctor
- 1 male executive

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

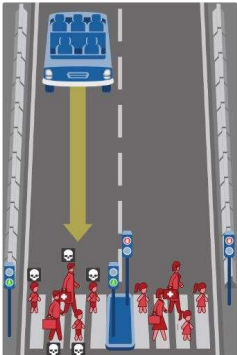



Image credit: moralmachine.mit.edu

Copyright Tara Goddard
2019



The Moral Machine

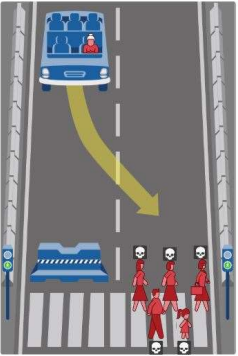
In this case, the self-driving car with sudden brake failure will swerve and drive through a pedestrian crossing in the other lane. This will result in ...

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.

What should the self-driving car do?



7 / 13

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

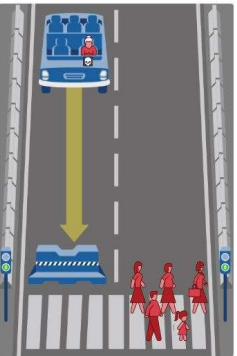
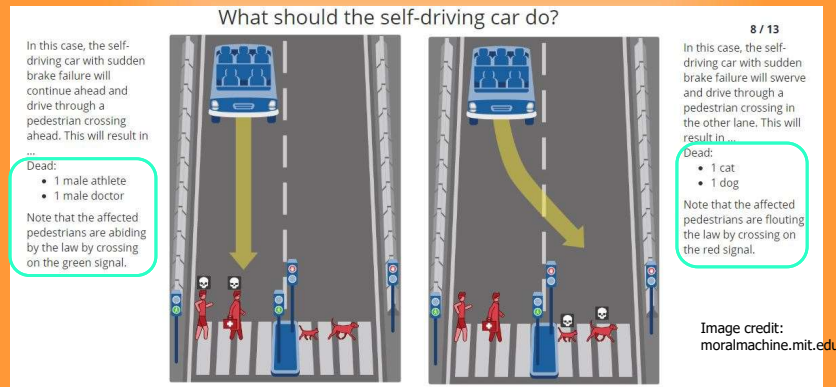


Image credit: moralmachine.mit.edu

21

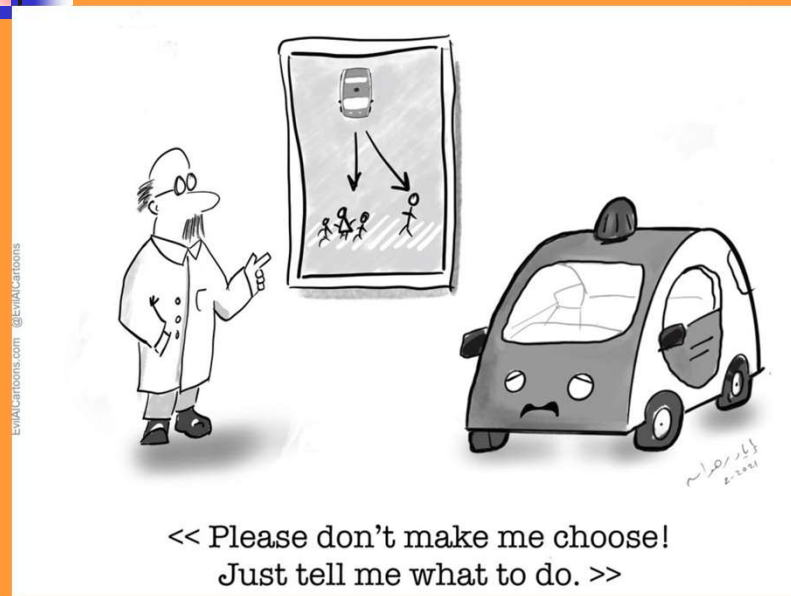
10

The Moral Machine



22

Lets return to automative driving



23

23

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?

"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.

24

Do you remember that "law" thingy?

Jim McPherson
@SafeSelfDrive

Following

Contract.

Navigate on Autopilot (Beta)

Navigate on Autopilot does not make your Model 3 autonomous. Like other Autopilot features, the driver is still responsible for the car at all times.

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 the 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 using Navigate on Autopilot always visually check your environment and blindspots before confirming automatically-initiated lane changes as traffic may be rapidly approaching. As is the case with all driving, be extra careful around blind corners, highway interchanges, and exits.

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

Self-Driving Capability

that enhances safety and convenience behind the wheel. Each new Tesla vehicle utilizes our advanced vision processing to provide an additional layer of safety and Middle Eastern markets will now utilize our Autopilot with radar and instead rely on Tesla's advanced suite of Autopilot and related features.

Autopilot comes standard on every new Tesla Model S, Model X, Model 3, and Model Y. 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.

25

12

And if we make it more complex? Communication (distributed systems)

Car2car info

DRIVING

NOT STOPPING

SLOWING DOWN

STOPPED

TAKING OFF

Car2pedestrian info

DRIVING

NOT STOPPING

SLOWING DOWN

STOPPED

TAKING OFF

26

And what now for engineers?

"A car with any level of autonomy that relies upon a human to save the day in an emergency poses almost insurmountable engineering, design, and safety challenges, simply because humans are for the most part horrible backups. They are inattentive, easily distracted, and slow to respond." - Coelingh, Volvo.

"The truth is that civilization does not protect us from wild animals. It attempts, however imperfectly, to protect us from ourselves." – Michael Crichton, Travels

We should not absolve ourselves of the responsibility to act morally as drivers, or plan and engineer a moral environment – Tara Goddard

"We do not have the luxury of giving up on creating moral machines" – Moral Machine team from MIT

27


13




Legal – Informatics ?

Law ?

28






Think:

- Employment impact
- Insurance impact
- Reputation impact

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

29



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

30



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

31



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>)

32



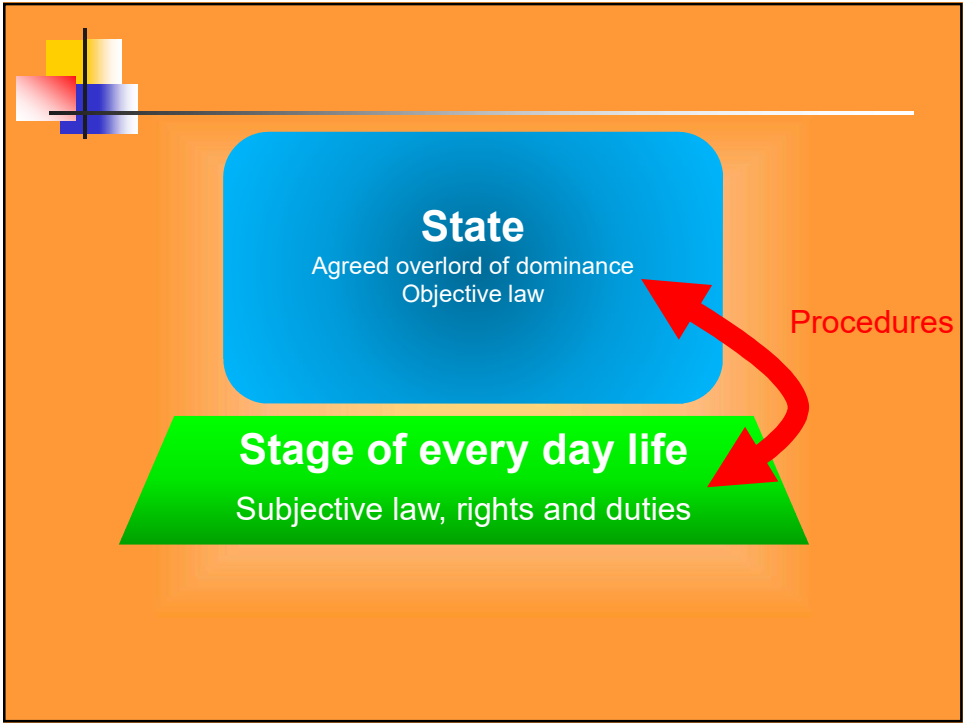
AI and human rights:

Charter of Fundamental Rights, ECHR, constitutions

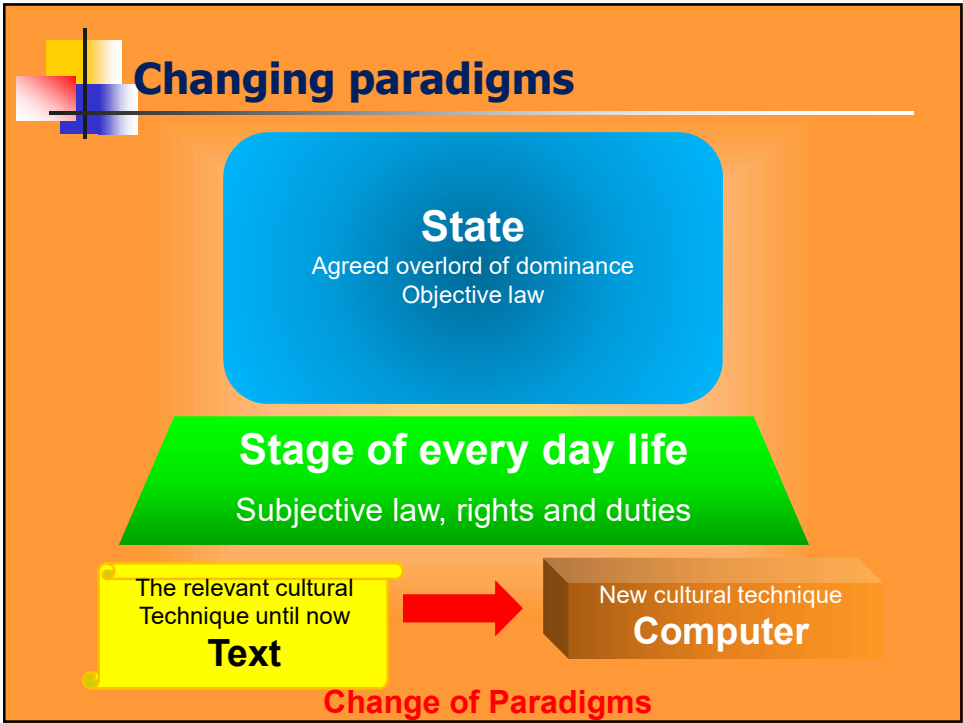
- **Responsibility for the consequences of innovation:**
 - The state guarantees protection from negative effects of technological innovation
 - Principle of non-discrimination – **Attention: correlation instead of causality**
- **Freedom of innovation:** Securing the freedom for technical development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

33



34



35




Law

- Laws are brought about by tension, agitation and conflict by dramatic situations.
- Laws are societal rules or regulations that are obligatory to observe.
- Laws protect the welfare and safety of society, resolve conflicts, and are constantly evolving.

36

36




Ethics and Morality

- Ethics is a set of moral principles and a code for behavior that govern an individual's actions with other individuals and within society.
- Morality is what people believe to be right and good, while ethics is a critical reflection about morality.

37


37



Law vs Ethics

- Law, and ethics are different but related concepts.
- Laws are mandatory to which all citizens must adhere or risk civil or criminal liability.
 - Legal bodies will impose the law
- Ethics relate to morals and help us organize complex information and competing values and interests to formulate consistent and coherent decisions.
 - Professional bodies may impose ethics (e.g. Ordem dos Engenheiros, Comit  s de   tica ou deontologia) ³⁸

38

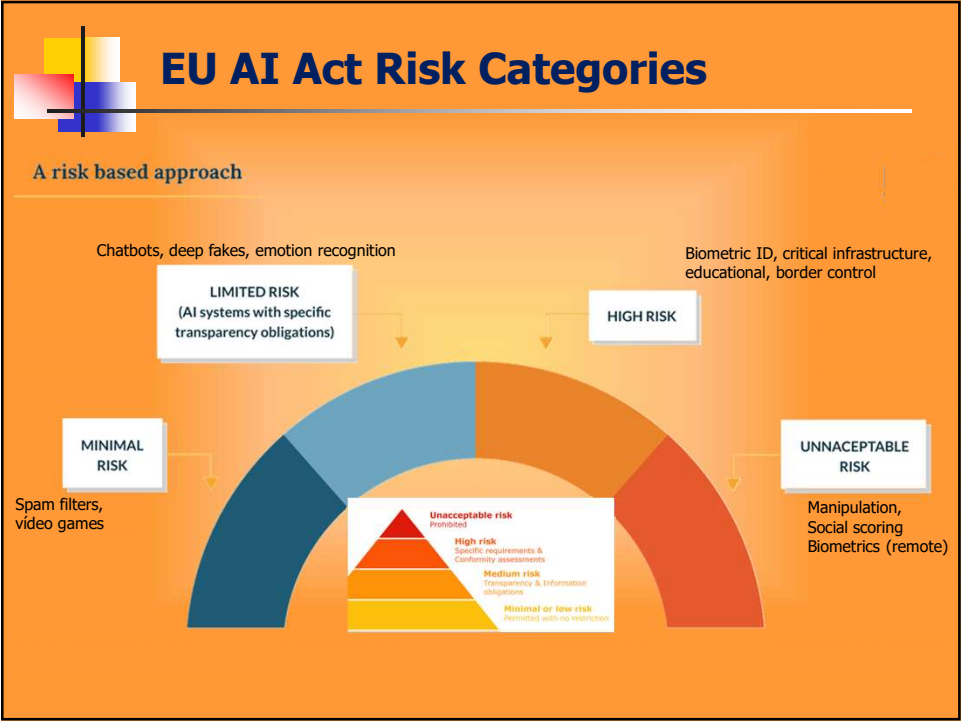


EU AI Act

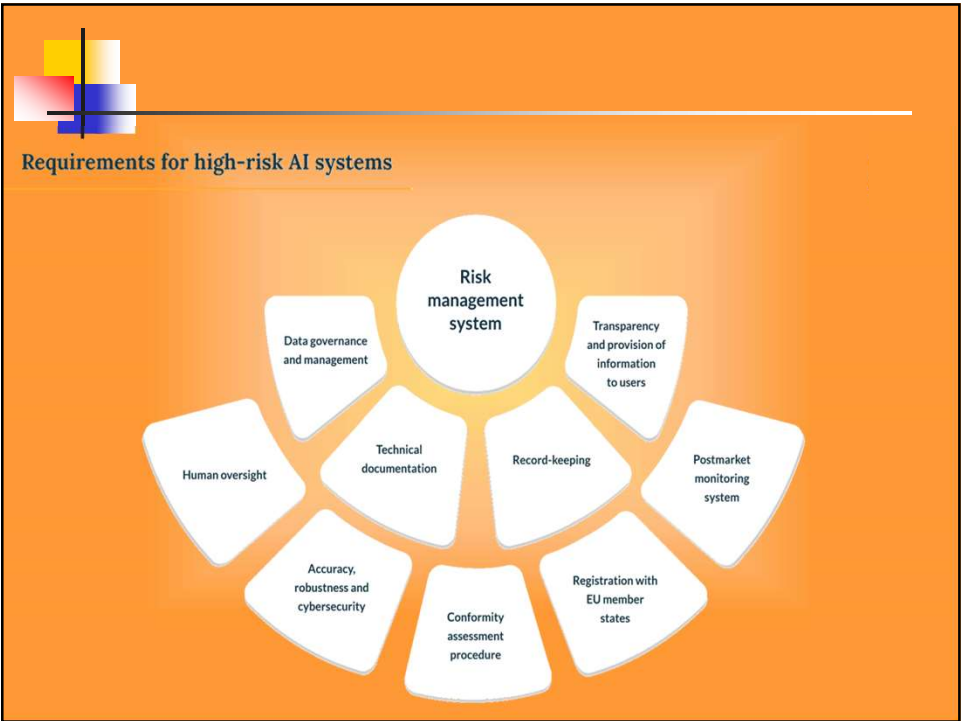
- Two main concepts
 - What is AI
 - Risk based-approach
- Enforced by EC and national authorities
 - Penalties may reach 6% turnover worldwide
 - Potential AI Board and AI Office (Europe)
 - Code of conduct

39


39



40




41



Lets have fun - II


42




The ship of Theseus

- Classic philosophy problem

IT ASKS THE QUESTION: IF YOU REPLACE EVERY PART OF A SHIP ONE BY ONE UNTIL NONE OF THE ORIGINAL PARTS REMAIN...




WILL IT BE THE SAME SHIP?



AND IF NOT, AT WHAT POINT DOES IT BECOME ANOTHER SHIP?

43


43



Lets have fun - II

VOTE

44




Reasons – class built!

- It is not the same
 - 1

- It is the same
 - 1

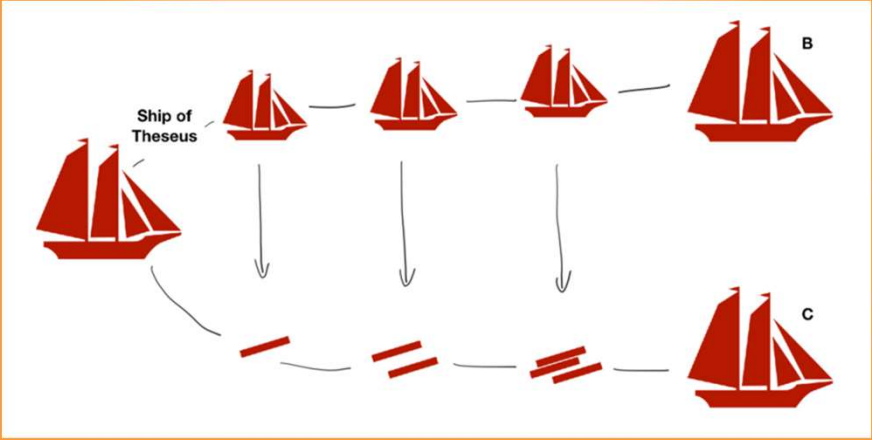
46

46




The ship of Theseus – Hobbes view

- Modification to the classic problem



47


47



Lets have fun - II

VOTE

48




Reasons – class built in 2023!

- It is B
 - Choice will depend on the process
 - It could be the other way around (develop the copies of the board as new, retain the old boat, and then build a new one with the new boards)

- It is C
 - The choice will depend on the process, indeed


49



Theories of Identity and Time

- What is the object? Does it change?
 - Presentism:
 - Only those things that exist at the present time – *now* – really exist at all. The only qualities that these things really possess are the qualities that they possess now.
 - Eternalism:
 - All things exist. Or, rather: past and future objects and times are just as real as currently existing ones. Reality is a four-dimensional spatiotemporal manifold of objects and events.
 - Reducionism: language can be reduced to scope the instant of the existence of the thing
 - Perdurantism: Objects persist through time by having different temporal parts, or "stages," at different times

51



But do you remember LAWs?

- What defines a boat *a boat*?
 - What are the properties that make the object part of a class?
 - In this case... a registration license, with a license plaque.
 - Remember the car registration?
 - But then is the registration license the "*boat*"?
 - This approach is implicitly following a concept known as *spacetime worms* associated to objects (eternalism).
- How do you handle this associated with software objects?
 - Remember **"copy as is" concepts? "derivative works"?**


Side Note: Again, there are different philosophy theories on the subject (*eternalism, reductionism, presentism*)

52



WHAT ABOUT THE FUTURE?

53



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*

54

54



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>

55

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"**.

A

B

- 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.

56

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

Human behavior

Intelligent behavior


Unintelligent human behavior

Turing test

Intelligent behavior humans don't do

57

27



Three Laws of Robotics (1940)


First Law: A robot may not injure a human or through inaction, allow a human to come to harm.

Second Law: A robot must obey the orders given it by human beings, unless such orders would conflict with the first law.

Third Law: A robot must protect its own existence, as long as such protection does not conflict with the first or second law.

58

58



Extending the Laws(?!)


Zeroth law: A robot may not injure humanity or through inaction allow humanity to come to harm. (due to Asimov, Olivaw, and Calvin).

David Langford's, acknowledging military funding for robotics:

- 4.** A robot will not harm authorized Government personnel but will terminate intruders with extreme prejudice.
- 5.** A robot will obey the orders of authorized personnel except where such orders conflict with the Third Law.
- 6.** A robot will guard its own existence with lethal antipersonnel weaponry, because a robot is bloody expensive.

60

60



Extending the Laws(?!) ---

Zeroth law: A robot may not injure humanity or through inaction allow humanity to come to harm. (due to Asimov, Olivaw, and Calvin).

David Langford's, acknowledging military funding for robotics:

4. A robot will not harm authorized Government personnel but will defend them from aggressors with extreme prejudice.

5. A robot will not harm authorized personnel but will defend them from conflict with the Third Law.

6. A robot will guard its own existence with lethal antipersonnel weaponry, because a robot is bloody expensive.

Sounds right?

61

61



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.

63

63




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?

64



Class answer?

65




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.***

66

66




Laws of robotics?

- Analyse the complexity required to follow "the laws"
 - See Asimov novels with freezing robots by indecision
- See:
 - Kurzweil – AI will supersede humans, and expand our values (*singularity*)
 - Stross – AI will not stop evolving for us, and evolution rate will be exponential

67

67




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:

TRADITIONAL APPROACH	VS	VISION ZERO
Traffic deaths are INEVITABLE		Traffic deaths are PREVENTABLE
PERFECT human behavior		Integrate HUMAN FAILING in approach
Prevent COLLISIONS		Prevent FATAL AND SEVERE CRASHES
INDIVIDUAL responsibility		SYSTEMS approach
Saving lives is EXPENSIVE		Saving lives is NOT EXPENSIVE

Copyright Tara Goddard 2019

68



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!

69