RJZR7

<https://www.moralmachine.net/>

Consequentiality theory: unclear with consequences are about. Consequentialism is not really a theory, depends a lot of different people believes and ability to calculate further down the line

focus on values. One of the theories withing consequentialism family is **utilitarianism**.

What you should do depends on the cosneuqeucneces of available actions. Theory of duty. Need a way to evaluate our actions. When the consequentialism emerged (late 1800) the evaluation criteria of the action was pleasure and pain (**hedonism**, particular type of consequentialism) ⟹ you want to pursue pleasure and avoid pain. This should guide our conduct.

Source of values: for whom? for the largest number? Still similar to consequentialism today. Hedonism is a primitive theory. Does not discriminate between the different kind of pleasure and pain (no focus about duration or intensity). For hedonism all the different pleasure are equal (same pleasure for reading a book or have sex) and all the pain are on the same scale (same pain of failing a test as a relative die). Didn’t go well with the present.

instead of judging the consequence you should act according to a principle that maximizes the consequences.

The action that is obligatory to perform is the one that follows the norms. But they way this norm is emplaced is because that norm has good consequences. Even if in that instance might be counter-intuitive.

Consequences are still involved.

What kind of values you might use.

Historically speaking it was pleasure and pain. Then someone else came up with the more technical notion of utility

Apple pie experiment[[1]](#footnote-0). The pleasure of the first apple pie is not the same of the fourth slice of apple pie. Same with monetary value. Suppose you can choose between a lottery ticket (A) with 1% change of winning 100million euros. Or alottery ticket (B) that gives 0.01% of winning a billion euros.

Not a normalitve stands. Psychologically speaking taking the higher change with a lot of money vs the one with lowest %. There is a decreasing value. The first slice of apple pie is great, but the 4th one does not give the same amount of pleasure or values.

The value of money/apple pie is not constant. You need to have an overarching theoretical value. First one valid 1 utility, the 4th one are 0,1 utility

What you have to do is to an action that maximises utility.

Act in such a way that your action furfill your duty or follow one principle. What is your duty?

deontological ethics

10 commandments give substationa advice and give you standard thighs to follow. **Deontological approach**. it is substatintail, just tell to not lie or not kill.

Kant: give a general recipty for action. Is up to you and your rational power how the conduct…

the (two) main problem of deontological theory:

1. **How to justify these principles**. Why these duty and not others. There are different ways to justify this (*theological*: god told me. Or by *Kant*: if you think rationally you arrived at the same end).

Find in many ways the golden rule. Do to others what you want the others to do to you. Example of a moral duty that is formal. Don’t tell what to do (kill everyone because you want to be killed).

1. **Moral dilemmas**. Syppose you should prevent any violent death and you should not lie. If someone nocks at your door and tells that they are followed by a killer and to hide him at your place. If you say yes and you hide him in the closed when the killer nocks. What tyou should do? you should not lie, and tell the truth, at the same time either you lie or you dont lie but both against your duty.

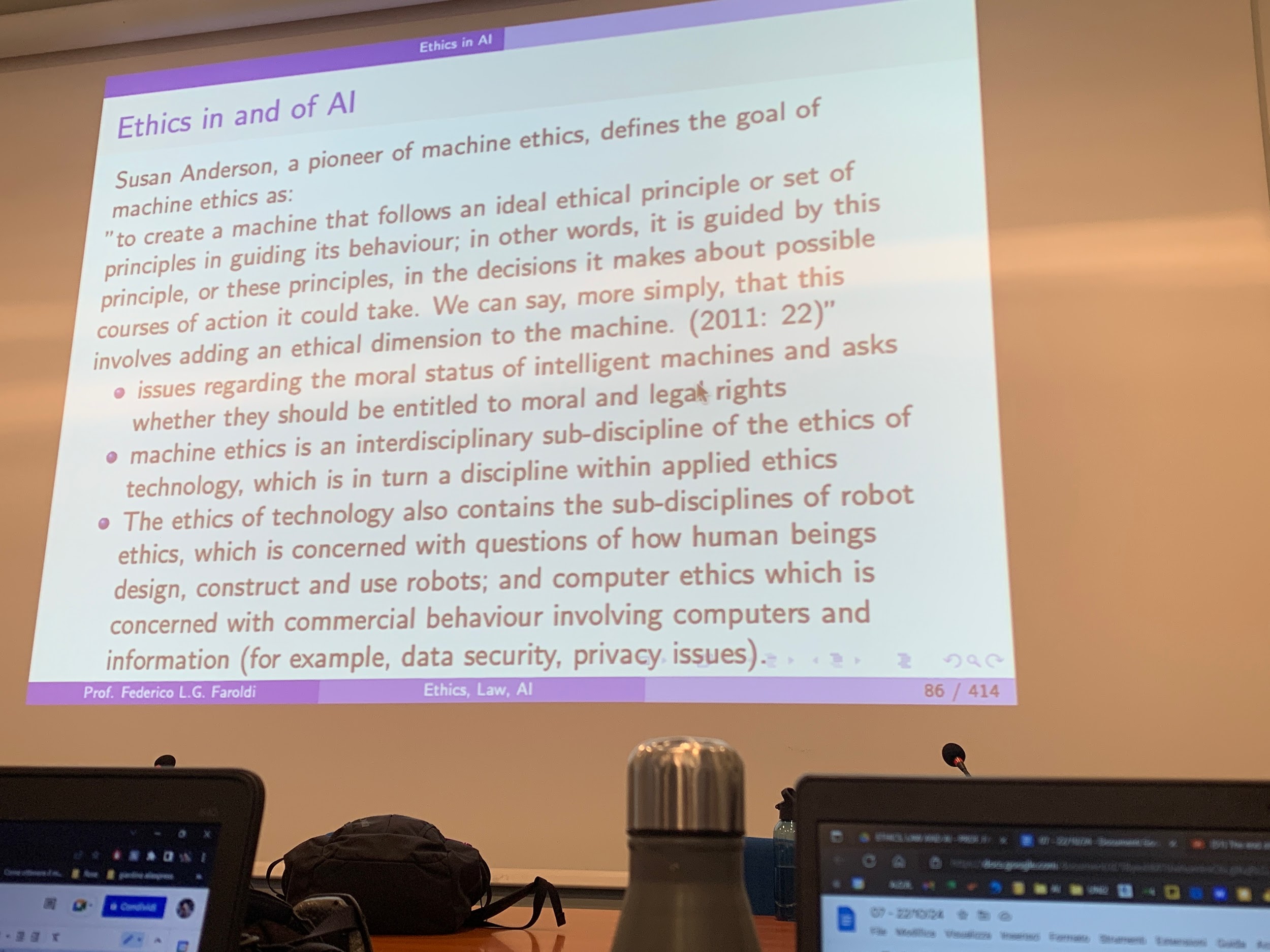
**State of necessity**: if you are in state of necessity (really starving for two weeks and you steal some food) it constitute a state of necessity. Can’t be found guilty of stealing.

**Existential risk**

Nick bostrom (ted talk)

[The end of humanity: Nick Bostrom at TEDxOxford](https://youtu.be/P0Nf3TcMiHo) (?)

machine etics:



How to be sure it behaves like it was ethical. Is an applied ethics issue, how we build ethical machines in a way that is etical to us. It also creates new metaethical question. We could ask if it is even possible that machines behaves ethically.

Might be subject of moral rights but questionable if non.human aniaml have moral duties. and can not be hold countable for the action they do in a moral way. Under one reading machine ethics is a subfiled of +compl

3 levels of metaetchis.

* Language (sematics and logical)
* Ontological (do valuers exist)
* Epistemologicay (how do we get to know them)

Suppose I tell *today is sunny*, is the question either true or false? yes. How you decide if it is true or false? by observing the reality. you use observation. empirical observation

if i ask if 394949389 is prime number is still a true or false question but you don’t anser by observing the reality.

abortion is wrong? can moral sentices be true or false? in order to determine if it ist rue or faslse first anser the ontholigcal question if there is a moral value.

at the same time you have to anser an epistemological question. can be empirical observation.

1. I give you a slide of apple pie (or something that you like) you are very happy. Then if I keep giving you again other slices the pleasure is not the same as the first slice (go decreasing) [↑](#footnote-ref-0)