RESPONSIBLE AI: LESSON 2

1. What is ethics and morals?

* Ethics: Moral principles that govern a person’s decision-making and behavior
* Morals: set of normative values, rules and ideals that are accepted in a society

1. Can and should AI make moral judgements?

* Humans are considered to be moral agents. Which is someone who has the ability to discern right from wrong based on their ethics, and be held accountable for their actions.
* However, in AI, moral situation are often not defined, thus the AI doesn’t know what to do
* Are AI really moral agents, when they must be made from different people with differing values/morals?

1. The moral responsibility test for AI

* A set of Aristotelian conditions to evaluate whether an AI has moral responsibility
  + Rule 1: cause an outcome through their own initiated and controlled (in)action (**causation condition**)
  + Rule 2: act physically and psychologically on their own will and intention (**freedom condition**)
  + be knowledgeable of the relevant details regarding the context of (in)action (**knowledge condition**)
  + possess the capacity to morally evaluate the significance of their action and inaction relative to a purpose (**deliberation condition**)

1. What is an ethical dilemma?

* decision-making problem between two possible moral imperatives, neither of which is unambiguously acceptable or preferable

1. The **Two Important Ethical Schools**:

* **Deontological ethics**: The morality of an action is based on whether that action itself is right or wrong under a series of rules. Some decisions/actions are considered morally obligatory regardless of their consequences
* **Utilitarian/Consequative ethics**: The consequences of one’s decisions are the ultimate basis for any judgment about the rightness or wrongness of that conduct. Moral rightness does not depend directly on anything other than effects. Outcomes can justify the means.
* Deontological Critic: By the statement, lying is always a bad thing. But, if lying would save a life, would it still be bad?
* Utilitarian Critic: If you use a flamethrower to do a good act, can the action still be morally right?

1. MIT Moral Machine

* each dilemma simultaneously randomizes three additional attributes:
  + which group of characters will be spared if the car does nothing;
  + whether the two groups are pedestrians, or whether one group is in the car;
  + and whether the pedestrian characters are crossing legally or illegally
* Global preferences:
  + • sparing humans over animals • sparing more lives • sparing young lives
* Local preferences:
  + For the West and the South, sparing the young is higher in priority than in the East
  + Preference for inaction is stronger in the West, and is less prominent in the South
  + South = southern america, West = europe/US, East = middle east / asia