Social Impact of Artificial Intelligence - Lecture 18.5

Topic:

- Ethics

Ethics - 1

- Moral and ethical issues pervade consideration of the impacts of AI.
- Morals are guidelines that apply to an individual's sense of right and wrong.
- Ethical principles apply at the level of a community, an organization, or a profession.
- Morals and ethics are connected: an individual may have morals derived from the ethics of groups they belong to.

Ethics - 2

- Ethical codes are categorized as either virtue-based, consequentialist, or deontological.
  - Virtue ethics emphasize the values and character traits of a virtuous agent.
  - Consequentialist (or utilitarian) ethics focus on the outcomes of possible actions that the agent can take, measuring the global utility of each outcome.
  - Deontological (or Kantian) ethical codes are based on a set of rules the agent should follow.

Ethics - 3

- AI ethics are motivated, in part, by worries about whether AI systems can be expected to behave properly.
- Q: Can we trust AI systems? A: Not now.
- Can they do the right thing? Will they do the right thing?
- Trust is not just about the system doing the right thing.
- System seen as trustworthy *only* if confident that it will *reliably* do the right thing.

Ethics - 4

- See movies and books: we fear that robots, and other AI systems, are untrustworthy.
- Could they become completely autonomous, with free will, intelligence, and consciousness?
- Will they may rebel against us as Frankenstein-like monsters?
- Trust issues lead to ethics concerns.
- If AI designers, deployers, and users are following explicit ethical codes, those systems are more likely to be trusted.
- If AI systems actually embody explicit ethical codes, they are also more likely to be trusted.

Ethics - 5

- What are ethical activities for developers of AI systems?
- For deployers of AI systems, are there applications that should not be considered?
- Should humans be guided by ethical principles when interacting with AI systems?
- Should AI systems be guided by ethical principles, in their interactions with humans or other agents?
- What data should be used to train AI systems?
- For each of these concerns, who determines, and enforces, the ethical codes that apply?

Ethics - 6

- AI ethics addresses two, distinct but related, topics:
  A. AI ethics for humans: researchers, designers, developers, deployers, and users
  B. AI ethics for systems: software agents and embodied robots
- Each topic considers ethical codes, of one of the three code types, either for humans or for systems.
- For topic A, need for professional code of ethics for AI designers and developers.
- Engineering, legal, medical, and computing professions all have explicit deontological ethics codes.
- For computing, the ACM, AAAI, and IEEE have ethics codes for members.

Ethics - 7

- What ethical issues arise for us, as humans, as we interact with AI systems?
- Should we give them any rights?
- There are human rights codes. Will there be AI systems rights codes, as well?
- Distinguish among moral agents, moral patients, and other agents.
- Moral agents can tell right from wrong and can be held responsible for their actions.
- A moral patient should be treated with moral principles by a moral agent.

Ethics - 8

- A typical adult human is a moral agent, and a moral patient.
- A baby is a moral patient but not a moral agent.
- A (traditional) car is neither.
- Could an AI agent ever be considered a moral agent? Should it ever be?
- Should current AI systems be considered as moral patients, warranting careful ethical treatment by humans? No?
- Could future AI systems, including robots ever be treated as moral patients? Should they be?

Ethics - 9

- See case study (Section 18.7) on the ethics of four forms of facial recognition, each with their own risks and benefits:
  - facial detection finds the location of faces in images.
  - facial characterization finds features of individual faces, such as approximate age, emotions.
  - facial verification determines whether the person matches a single template.
  - facial identification is used to identify each person in an image from a database of faces.
- Facial identification, usually considered the most problematic, has difficulties that arise both when it is perfect and when it makes mistakes.

Ethics - 10

- Consider AI ethics for systems. How should AI systems make ethical decisions as they develop more autonomy?
- The Laws of Robotics [Asimov, 1950] impose limitations on robotic behavior. Asimov's original three laws are:
  I. A robot may not harm a human being, or, through inaction, allow a human being to come to harm.
  II. A robot must obey the orders given to it by human beings except where such orders would conflict with the First Law.
  III. A robot must protect its own existence, as long as such protection does not conflict with the First or Second Laws.

Ethics - 11

- The prioritized laws of robotics should be followed by all robots and, by statute, manufacturers would have to guarantee that.
- They constitute a deontological code of ethics for robots, imposing safety constraints on acceptable robotic behavior.
- Discussions of AI ethics for systems, often presuppose technical abilities to impose and verify AI system safety requirements that just do not exist yet.
- But some progress on formal hardware and software verification is described in Section 5.10.