INFO 350

Autonomous Agency and AI Ethics

Algorithms & Autonomous Agency

And robots!

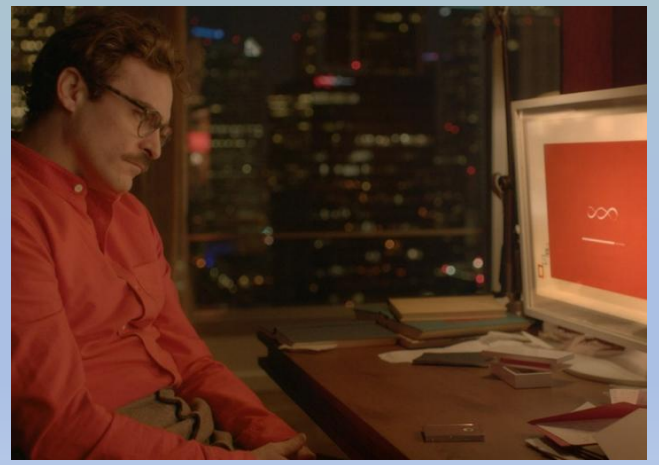
Should Robots Have Rights?

Should Robots Have Rights

- Humans are prone to assigning ”sentience” to objects
  - Naming cars, bicycles, etc.
  - Believing computers have feelings, feel your pain when frustrated, etc.
  - Proposing marriage to Siri (happens a LOT)
  - Believing anything that moves can also think and feel
- We call this “anthropomorphism”

Pressing Questions

- Are we getting to the point where ”intelligent” devices resemble “life?”
- If we are not there yet, will that day come? When?



Alan Turing and “The Imitation Game”

- Also known as “The Turing Test”
  - Designed to answer the question “can machines think?”
  - Alan Turing (1950) argued that if a machine can respond in a way that makes it impossible to distinguish it from a human, we can say it *thinks*
- The Turing Test was a thought experiment – it was not attempted on computers of Turing’s time

Saudi Arabia becomes first country to grant a robot citizenship – and people are saying it already has more rights than women

Unlike Saudi women Sophia the robot appeared alone without a male guardian

Natasha Salmon | Saturday 28 October 2017 16:50 | | 29 comments

Should Robots be Slaves?

Bryson:

- Robots should not be labeled or treated as persons
- They should not bear moral or legal responsibility for action
- They should be considered property – owned by “us”
- Humanizing robots dehumanizes people

- Robots should be slaves

Robots should be slaves?

- This is a pretty inflammatory statement.
- Critics have ”pounced” on this.
- What criticisms do you have?

What is a robot?

- “A robot is any artificial entity situated in the real world that transforms perception into action. “
  - Not exclusively human or animal like devices
  - Do not have to be capable of acting – as long as they can cause action
  - Siri, Alexa, Cortana – are robots
  - “Por español, marque dos” – phone bots are robots
  - Bots = robots!

Agency

- **Agency:** The ability to act as a rational, moral being. Also, the ability to *choose* from among possible actions. Also, ability to be held *accountable* for action.
  - Aka “moral agency”
- **Agency** creates obligations on others: we must *treat* agents as deserving of moral consideration (as ends, not means…)
- Bryson: Robots do not have agency even if they appear to.
  - They ought not be considered “friends”
  - Their interests are subservient to ours

Efficiency

- P1: We have a limited amount of time, money, other resources
- P2: Resources expended on human-robot interactions are resources taken from human-human interaction
- P3: Human-human relationships are fulfilling (even if complex) than human-robot relationships.
- P4: Human-robot relationships are less fulfilling (even if easier) or not at all fulfilling compared to human-human relationships
- C: Therefore, resources should prioritize human-human relationships

“one should spend money and time on [robots] as is appropriate to their utility, but not much more.”

Inefficiency

Convenience and Liability

- Some argue that robots make “more ethical decisions” than humans, and can operate without fatigue and at lower cost.
- Bryson counters: we already see automated agents making decisions and providing services but do they really substitute for humanity?
  - Railway ticket kiosk example
- Robots make decisions that reflect the priorities of their owners and designers. Granting robots (what appears to be) decisional autonomy is just a way to shift corporate responsibility onto something else.

What do we owe robots?

- Robots can/will only have the experiences we provide to them.
  - Can a robot feel “frustration?”
  - Can a robot suffer?
  - Can a robot die?

- “Robot builders *are* ethically obliged — obliged to make robots that robot owners have no ethical obligations to”
  - Robots are tools
  - “we are obliged not to the robots, but to our society.”

Autonomous Agency - Lethal

Autonomous Weapons Systems (aka ”killer robots”)

Point/Counterpoint

About half of you will read the case *against* (Goose)

About half of you will read the case *in favor* (Arkin)

Artificial Intelligence

Regulating Artificial Intelligence

- Artificial intelligence (AI) is playing an increasingly important role in human lives
- We have some options:
  - Promote speedy, “permissionless” innovation and deal with problems as we encounter them
  - Prioritize AI safety and other considerations through regulation even if it slows down innovation
  - Decide not to develop or use AI in most or all cases

What path would YOU choose?

When poll is active, respond at **PollEv.com/mikekatell776** Text **MIKEKATELL776** to **22333** once to join

Promote speedy innovation, regulate or fix things when they come up

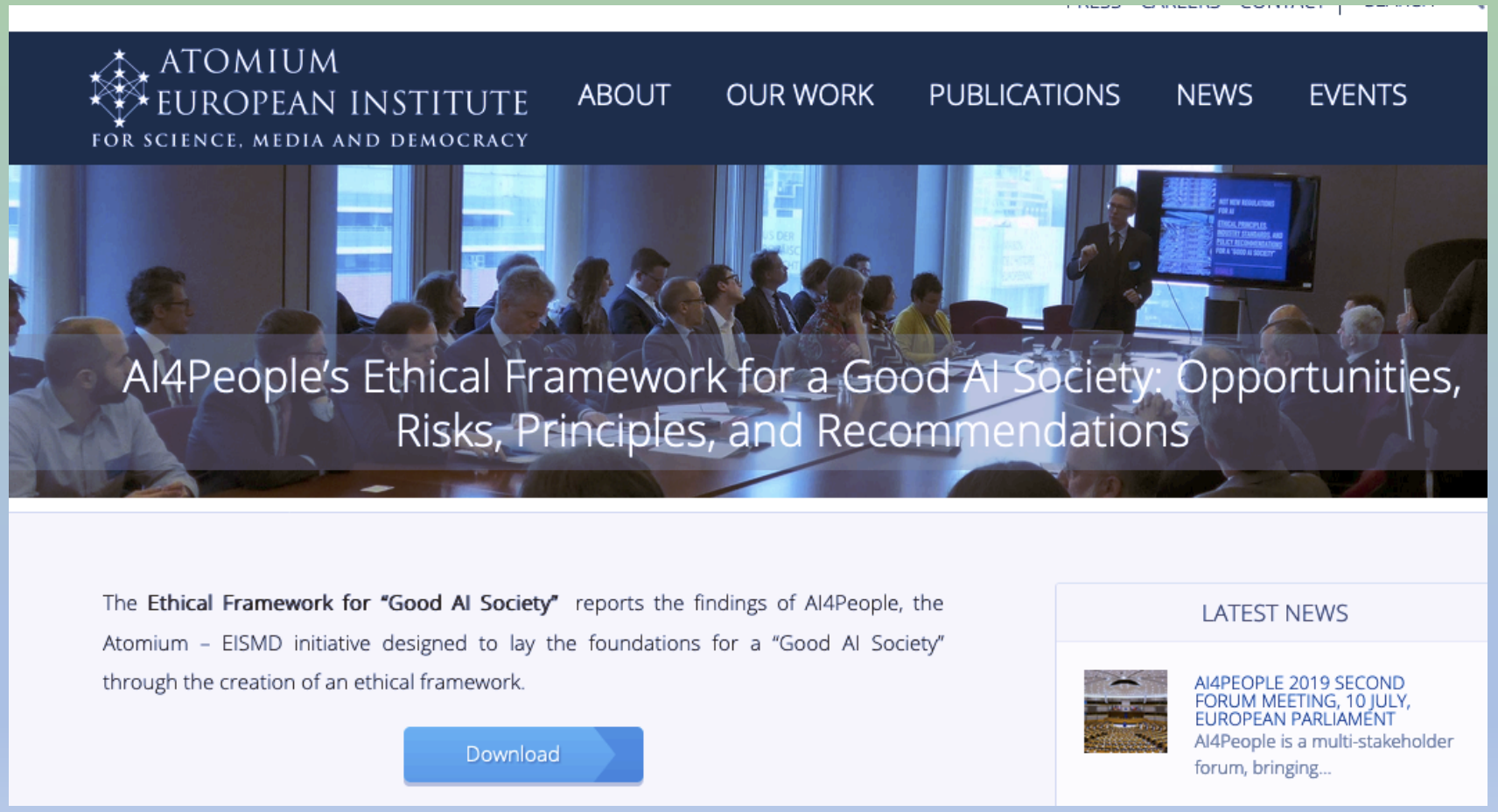
Promote safety and other considerations above innovation

Decide not to pursue some or all artificial intelligence because it's too risky

Do something else…

AI Ethics - frameworks

AI Ethics - frameworks



Agency

- AI could increase human agency – giving humans new abilities to control the world, their fates, etc.
- AI may create ways for people to have *less* agency. Responsibility can be deferred to AI (similar point to Bryson's)
  - ”Black box” mentality – AI decisions too complex for human understanding, so we just “go with it”

AI4People

AI4People proposes “A Unified Framework of Principles for AI in Society”

- Adapted and informed from review of existing tech/AI ethics principles by several organizations
  - Arrived at five key principles
  - Overlap and commonality with four core principles of *bioethics*
- Poses the question: are we ”the patient” or ”the doctor” where AI is concerned?
  - What does this mean?

Five Principles

- **Beneficence: Promoting Well-Being, Preserving Dignity, and Sustaining the Planet**
  - Well being of humanity, well-being of the planet
- **Non-maleficence: Privacy, Security and "Capability Caution"**
  - Intentional and unintentional harms should be avoided
- **Autonomy: The Power to Decide (Whether to Decide)**
  - Promote *human* autonomy.
  - Humans get to *choose* how and when to cede authority to AI
- **Justice: Promoting Prosperity and Preserving Solidarity**
  - Preventing discrimination, promoting solidarity
  - Righting of prior wrongs (and not creating new wrongs)

Five Principles

- **Explicability: Enabling the Other Principles Through Intelligibility and Accountability**
  - This one is *not* from bioethics – specific to AI and emerging tech
  - AI should be accountable and explainable (intelligible, transparent…)
    - Reasonable answers to “why” questions after AI acts