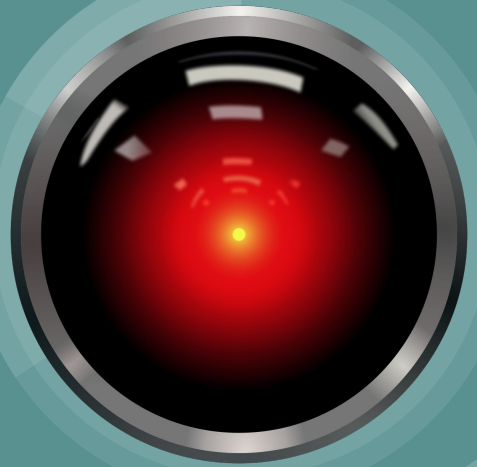
AI, Robots, and Ethics



“I’m sorry Dave, I’m afraid I can’t do that”

Welcome to the future: Big Data and Bigger Questions

Can a pigeon feel pigeonholed?

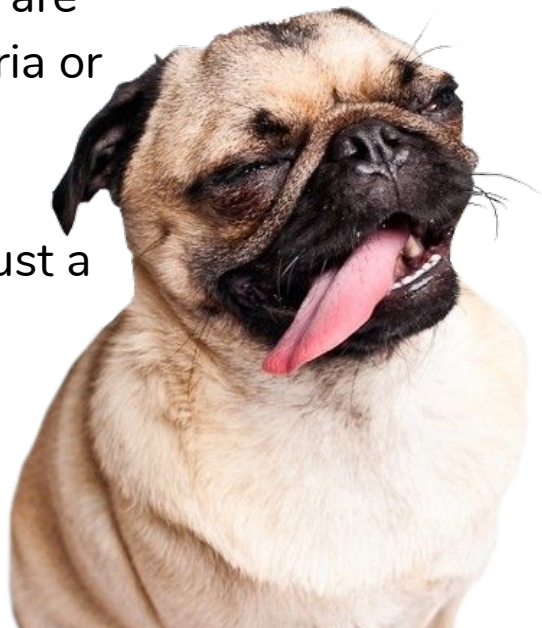
- Descartes - Doctrine of Bete-Machine
  - Animals and mechanisms are one in the same
  - Animals can be moral patients but not agents

What is agency?

Agency: “The idea of agency is conceptually associated with the idea of being capable of doing something that counts as an act or action…Actions are doings, but not every doing is an action; breathing is something we do, but it does not count as an action. Typing these words is an action, and it is in virtue of my ability to do this kind of thing that, as a conceptual matter, I am an agent.” (Himma 2009, 19-20).

Moral Agency?

- Includes only those agents whose actions are directed by or subject to some moral criteria or stipulation
- “If the dog brings in porn flyers from the mailbox and gives them to your kids, it’s just a dog, and it doesn’t know any better. If the butler does it, he is a legitimate target of blame” (J. Storrs Hall, 2007)



How do I know you even have a moral compass?

- You don’t! The other minds problem
- Sullin’s Moral Agency via mindless morality
  - Autonomy - not under direct control of another agent
  - Intentionality - Just have to ascribe beliefs/desires
  - Responsibility - Appearance is all that matters

Who or what is deserving of ethical consideration?

Animal	Machine	Man (Person)
● Moral patients ● Moral agents: a very exclusive list of higher-functioning mammals	● Machine vs. Tool ● The jury is still out	● OED: A self conscious or rational being ● We get all the rights

What arguments have opened the floodgates?

The Cumulative Argument for Animal Consciousness (Regan, 1983)

1. Commonsense view of the world.
2. Linguistic habits that demonstrate a conscious mind state
3. Humans aren’t special
4. Animal Behavior, which appears to be consciously directed and not random
5. Evolution!

That’s all great, but what happens while we figure out the answers?

Functionalist route (Applications route)

Asimov’s Laws of Robotics

1. A robot may not injure a human being or, through inaction, allow a human to come to harm
2. A robot must obey any orders given to it by human beings, except where such orders would conflict with the first law
3. A robot must protect its own existence as long as such protection does not conflict with the first or second laws

So if that’s the future, what about now?

What are algorithms?

“The math computers use to decide stuff”

-Slavin

How do we use algorithms?

- Black Box trading
  - Millions of micro-transactions
  - 70% of the US stock trading system
- Netflix
  - Determines 60% of the content people choose to watch

LogoJoy

PLUCK
PLUCK
PLUCK
PLUCK

1W
2W
1M
3M
1Y
ALL
GROSS VOLUME
$70,649.25
NEW CUSTOMERS
0
SUCCESSFUL PAYMENTS
1,259
$4.4K
$2.7K
$4.4K
$2.7K
02/05
02/09
02/13
02/17
02/21
02/25
03/01
03/06

Chef Watson

2050

That’s the year in which artificial intelligence will be able to perform **any intellectual task a human can perform**, according to James Barrat and Ben Goertzel. Anything and everything any person has ever done in all of history—all of it doable, by 2050, by intelligent machines.