When should humans overrule AI?

Dhileas Heywood
2020/3/16

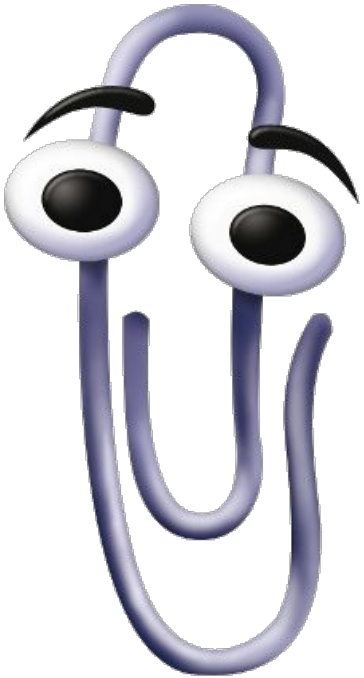
What is Artificial Intelligence?

- Machine Learning - Computational routines that predict future data from historical data.
- Deep Learning - A subset of Machine Learning. Computational routines that interpret data from pictures, videos, or other similar data types.

What is Artificial Intelligence?

A program that augments or mimics human cognition

Why do we need to control AI?



Clippy the paperclip from Microsoft Word

How can we control AI?

- A big red button?

Cooperative Inverse Reinforcement Learning

Dylan Hadfield-Menell* **Anca Dragan** **Pieter Abbeel** **Stuart Russell**

Electrical Engineering and Computer Science
University of California at Berkeley
Berkeley, CA 94709


But Wait!

What if the Human is Wrong?

Boeing 737 MAX Crashes

- Pilots passed the take off stage of flight without problem.
- Started to commence the ascent.
- When they started the ascent, the plane went into a nosedive.
- Sirens went off, alerting the pilots that they were too close to the ground, and warnings displayed that altitude and speed and pitch sensors were off.
- Pilots had to figure out why the plane was diving, as well as stop it.
- In two out of three malfunction cases, the plane crashed, killing everyone

Boeing 737 MAX Crashes

- Due to a failure in the Maneuvering Characteristics Augmentation System (MCAS)
- Due to a series of human errors.
  - Deciding to use readings from just one sensor to activate MCAS, instead of the original two.
  - Not disclosing the existence of the system to pilots before the first crash.
  - Not testing the system in the air, only in simulations.
  - Not submitting considerable changes to the system for review by the Federal Aviation Administration.

In conclusion

When should we overrule AI?

- Whenever it malfunctions.
- But:
  - We have to give humans the tools they need in order to overrule
  - We have to put measures in place to make sure the AI will allow it
  - We have to test AIs rigorously, not just the minimum required