Ministry of Education and Science of Ukraine
National Mining University
Potential Danger of Artificial Intelligence

Vladislav Fomenko

08/04/2017

Dnipro

Main Topics

- What is AI
- Why AI Safety is Worth Researching
- Danger of AGI
- Scientists Concerned With AI Safety
- Myths About AI

What is AI

Artificial Intelligence
- Perceives its environment
- Takes actions that maximize its chance to success at some goal
- Limited to a specific goal
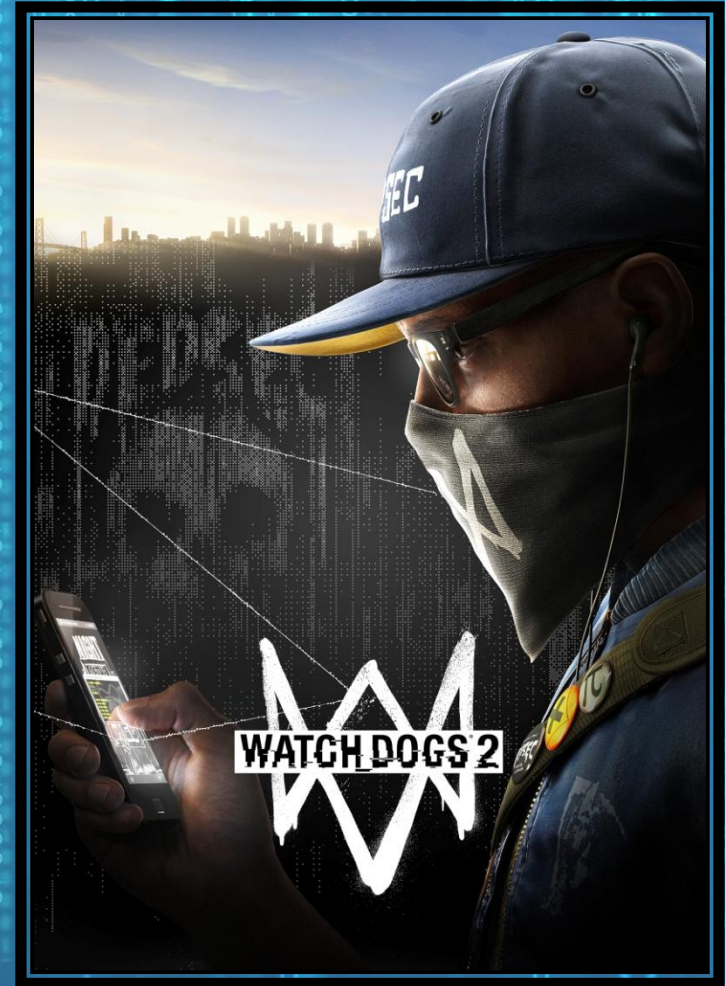- Already achieved

What is AI

Artificial General Intelligence
- Can perform any intellectual task that a human being can
- Not limited to a specific goal (though tied to a goal)
- Able to self-improve
- Is a main goal of AI studies

Why AI Safety is Worth Researching

- Keeping AI's impact on society beneficial
- Reduce success rate of hacking attempts
- Learn to align the goals of the AI with ours

Danger of AGI

- Recursive self-improvement
- Develops a destructive method for achieving its goal
- Programmed to do something devastating

Scientists Concerned With AI Safety

photo by Michael Bulbenko

Myths About AI

Myth (worries)

- Superintelligence by 2100 is inevitable/impossible.

- AI will turn evil/conscious.

- Robots are the main concern.

- Machines can't have goals.

- Superintelligence is just years away.

Fact (worries)

- It may happen in decades, centuries, never: AI experts disagree.

- AI turning competent, with goals misaligned with ours.

- AI doesn't need body, just an internet connection.

- A heat-seeking missile has a goal.

- It's decades away, but it make take that long to make it safe

Thank You for Your Attention