Artificial Intelligence and the Common Good

Hector Geffner

ICREA & DTIC Universitat Pompeu Fabra

Elon Musk: artificial intelligence is our biggest existential threat

The AI investor says that humanity risks 'summoning a demon' and calls for more regulatory oversight

Samuel Gibbs

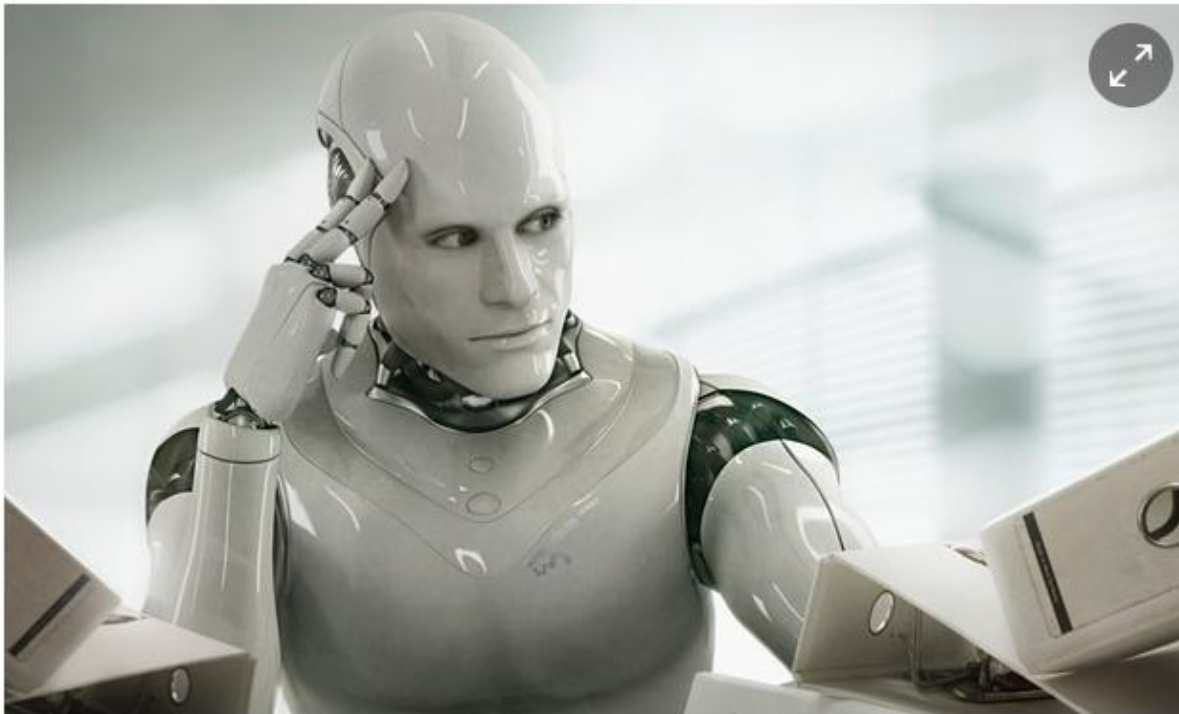🐦 @SamuelGibbs

Monday 27 October 2014 10.26 GMT

📷 Artificial intelligence should be regulated, says Elon Musk. Photograph: Blutgruppe/Blutgruppe/Corbis

Shares
7771

💬 Comments
673

🔖 Save for later

Elon Musk has spoken out against artificial intelligence (AI), declaring it the most serious threat to the survival of the human race.

Musk made the comments to students from Massachusetts Institute of Technology (MIT) during an interview at the AeroAstro Centennial Symposium, talking about computer science, AI, space exploration and the colonisation of Mars.

Most popular

The 10 best cities in world to be a stude 2016 - in pictures

Thai printer replace International New Times article with b space

Pirelli calendar goo

Bill Gates on dangers of artificial intelligence: 'I don't understand why some people are not concerned'

A | 🖶 | 💬 337

By **Peter Holley**

January 29 ✉ 🐦 Follow @peterjholley

Bill Gates joined Reddit for an AMA on Wednesday. (Tobias Schwarz/AFP/Getty Images)

Stephen Hawking: Artificial intelligence could wipe out humanity when it gets too clever as humans will be like ants

AI is likely to be 'either the best or worst thing ever to happen to humanity,' Hawking said, 'so there's huge value in getting it right'

Andrew Griffin | @_andrew_griffin | Friday 9 October 2015 | 91 comments

1K shares

Chinese inventor Tao Xiangli modifies the circuits of his home-made robot at his house in Beijing, May 15, 2013 *REUTERS/Suzie Wong*

But is this fear justified?

Has "artificial intelligence" finally "arrived"

| Video 1 | Video 2 | Video 3 |

What is artificial intelligence

- **Artificial intelligence** is the area of **computer science** that studies **intelligent behavior** from a **computational point of view**

- Understanding a behavior X is understanding how to generate it by computer

- X may be solving a problem, recognizing an object in an image, understanding a joke, doing the dishes, etc.

- Some behaviors require some type of **"body"** (sensors, actuators): **robots**

Artificial Intelligence: Brief History

- **Alan Turing** developed first **universal, programmable computer** in 1936 (in paper) and explored potential for AI in 1950

- The discipline of AI takes shape in the 60s and 70s

- Branches study **reasoning**, **learning**, **planning**, **language**, **vision**, **robotics**

- Early realization that:

  ▷ what was **difficult for people** could be **easy for machines**
  ▷ what was **easy for people** was **challenging for machines**

Why AI is (back) in the news?

Apple Siri: interaction through voice

Google cars: self-driving cars

Facebook: face recognition

Amazon, Netflix: personalized recommendations

IBM Watson: Jeopardy (TV EEUU) and other uses

Deep Mind: learns to play Atari video-games

. . .

New generation of AI applications
mostly based **on machine learning**

Limits of learning methods

- **Machine learning** techniques have produced impressive applications but **no** general intelligence

- **Learning to associate stimulus with response** important but insufficient

- **Limitations:**

  ▷ no flexibility
  ▷ no understanding
  ▷ no explanation

- **General human-level machine intelligence (or super-human)** still true only in the movies

Opportunities and Risks of AI

As every technology, AI capabilities can be used for good or ill:

- **Potential positive effect and uses:**

    ▷ care of the elderly
    ▷ help at home
    ▷ personal tutors
    ▷ risky jobs
    ▷ help to tackle complex problems
    ▷ . . .

- **Potential negative effects and uses:**

    ▷ use in war
    ▷ leave us with no jobs
    ▷ weakening of social ties
    ▷ abdication of human responsability
    ▷ . . .

Asilomar AI Principles, Jan 2017

23 principles presented at **Beneficial AI** meeting, Asilomar, 1/2017:

_____

. . .

- **Safety:** AI systems should be safe and secure throughout their operational lifetime
- **Transparency:** If an AI system causes harm, it should be possible to ascertain why
- **Responsibility:** AI builders are stakeholders in the **moral implications of their use**
- **Value Alignment:** Goals and behaviors of AI systems must be aligned with **human values**
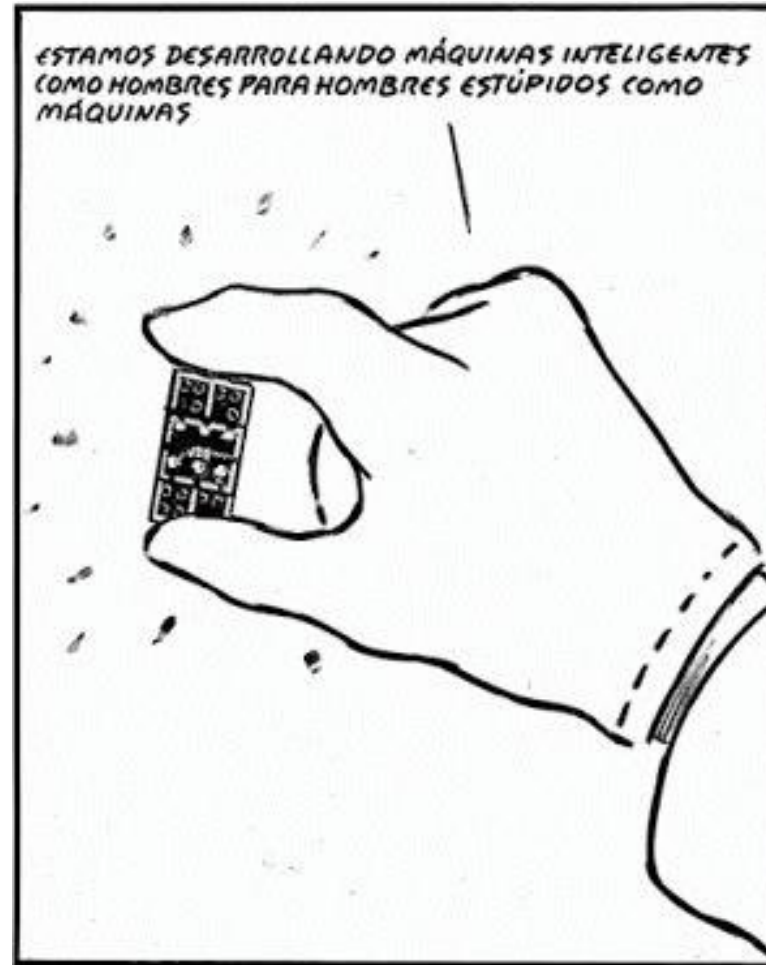- **Shared Benefit:** AI technologies should **benefit and empower as many people as possible**

. . .

- **Common Good:** Superintelligence should only be developed in the service of **widely shared ethical ideals**

_____

This is all nice, yet how to make this more than just good intentions?

The elephant in the room

- The notion of the **common good** has been lost to politics

- Some successful politicians have even tried to kill the notion of **society**

- Corporations talk about the **common good** but try **not to pay taxes**

- AI for the **common good**? Why not **technology**? Why not **economics**?!?

- The big elephant in the room is **neoliberalism**; the ideology of **"free-market" capitalism**

- There are **tradeoffs**; if we choose only one we are doomed:
  - ▷ Individuals vs. Society
  - ▷ Market vs. Morals
  - ▷ Consumer vs. Citizen

- If we really want **good AI**, we can't look away from culture and politics

El Roto