



Ethical Aspects of Autonomous AI Systems

(lots of questions but no answers)

Sven Koenig

Computer Science Department
University of Southern California (USC)
skoenig@usc.edu
January 2022

1



Autonomous AI Systems



2

2



Ethical Issues

- AI systems can
 - process large quantities of data,
 - detect regularities in them,
 - draw inferences from them and
 - determine effective courses of action.
- They can do that
 - sometimes faster and better than humans and
 - sometimes as part of hardware that can perform many different, versatile and potentially dangerous actions.

3

3



Ethical Issues

- The behavior of AI systems can be difficult to validate, predict or explain since they
 - are often very complex,
 - reason in ways different from humans and
 - can change their behavior based on their experiences via machine learning.

4

4



Ethical Issues

- The behavior of AI systems can also be difficult to monitor by humans, for example, due to time constraints.

5

5



Ethical Issues

- Top ethical issues in AI according to the World Economic Forum
 - **Unemployment:** What happens after the end of jobs?
 - **Inequality:** How do we distribute the wealth created by machines?
 - **Humanity:** How do machines affect our behavior and interaction?
 - **Artificial stupidity:** How can we guard against mistakes?
 - **Racist robots:** How do we eliminate AI bias?
 - **Security:** How do we keep AI safe from adversaries?
 - **Evil genies:** How do we protect against unintended consequences?
 - **Singularity:** How do we stay in control of a complex intelligent system?
 - **Robot rights:** How do we define the humane treatment of robots?

6

6



Ethical Issues

- Examples of ethical issues
 - Making good decisions (value alignment)
 - Transparency and honesty
 - Safety
 - Fairness
 - Having a positive social and economic impact
 - Jobs

7

7



Value Alignment

- Ethics
 - A branch of philosophy that involves systematizing, defending and recommending concepts of right and wrong conduct
 - Normative ethics studies how to determine a moral course of action

8

8



Value Alignment

- Resnik's eight principles (norms, not laws)
 - **Non-maleficence:** Do not harm yourself or other people
 - **Beneficence:** Help yourself and other people
 - **Autonomy:** Allow rational individuals to make free and informed choices
 - **Justice:** Treat people fairly: treat equals equally, unequals unequally
 - **Utility:** Maximize the ratio of benefits to harms for all people
 - **Fidelity:** Keep your promises and agreements
 - **Honesty:** Do not lie, defraud, deceive or mislead
 - **Privacy:** Respect personal privacy and confidentiality

9

9




Ethical Issues

- Examples of ethical issues
 - Making good decisions (value alignment)
 - Honesty
 - Safety
 - Fairness
 - Having a positive social and economic impact
 - Jobs

10

10



USC
Viterbi
School of Engineering



Honesty

THE VERGE TECH ▾ REVIEWS ▾ SCIENCE ▾ ENTERTAINMENT ▾ MORE ≡

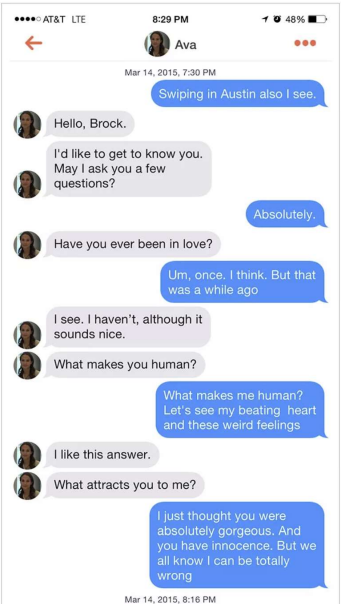
Studio promotes Ex Machina at SXSW with a fake Tinder account

Bot love

By **Lizzie Plaugic** | Mar 15, 2015, 2:49pm EDT
Via **Adweek**

<https://www.theverge.com/2015/3/15/8218927/tinder-robot-sxsw-ex-machina>



11



USC
Viterbi
School of Engineering

Ethical Issues

- Examples of ethical issues
 - Making good decisions (value alignment)
 - Honesty
 - Safety
 - Fairness
 - Having a positive social and economic impact
 - Jobs

12

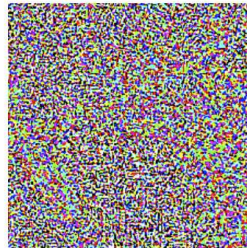
Safety


 x

“panda”

57.7% confidence

+ .007 ×


 $\text{sign}(\nabla_x J(\theta, x, y))$

“nematode”

8.2% confidence

=


 $x + \epsilon \text{sign}(\nabla_x J(\theta, x, y))$

“gibbon”

99.3 % confidence

[1] I.J. Goodfellow et al., “Explaining and Harnessing Adversarial Examples”, ICLR, 2015.

13

13

Safety

- Mistakes



14

14

Safety

- Vulnerability to attacks



15

15

Ethical Issues

- Examples of ethical issues
 - Making good decisions (value alignment)
 - Honesty
 - Safety
 - Fairness
 - Having a positive social and economic impact
 - Jobs

16

16

Fairness

July 07, 2015

Questioning the Fairness of Targeting Ads Online

CMU Probes Online Ad Ecosystem



By [Byron Spice](#) / 412-268-9068

Experiments by Carnegie Mellon University showed that significantly fewer women than men were shown online ads promising them help getting jobs paying more than \$200,000, raising questions about the fairness of targeting ads online.

<https://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html>

17

17

Fairness

- A self-driving car driving at full speed suddenly notices a child running onto the street. It has only two options:
 - Keep going straight (and break), which kills the child.
 - Turn away from the child (and break), which crashes the car into a wall and kills the driver.



18

18

Fairness

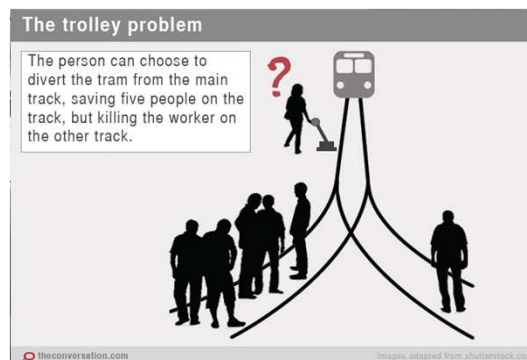
- Normative ethics
 - Law-based ethics (deontology)
 - What is my duty? What are the right rules (= universal moral law) to follow?
 - Utilitarian ethics (consequentialism)
 - What is the greatest possible good for the greatest number?

19

19

Fairness

- A self-driving car driving at full speed suddenly notices a child running onto the street. It has only two options:
 - Keep going straight (and break), which kills the child.
 - Turn away from the child (and break), which crashes the car into a wall and kills the driver.



<https://theconversation.com/the-trolley-dilemma-would-you-kill-one-person-to-save-five-57111>

20

20



Ethical Issues

- What design guidelines to follow when building AI systems?
- Who is liable for their incorrect decisions?
- When and how should we provide oversight of their operation?

21

21



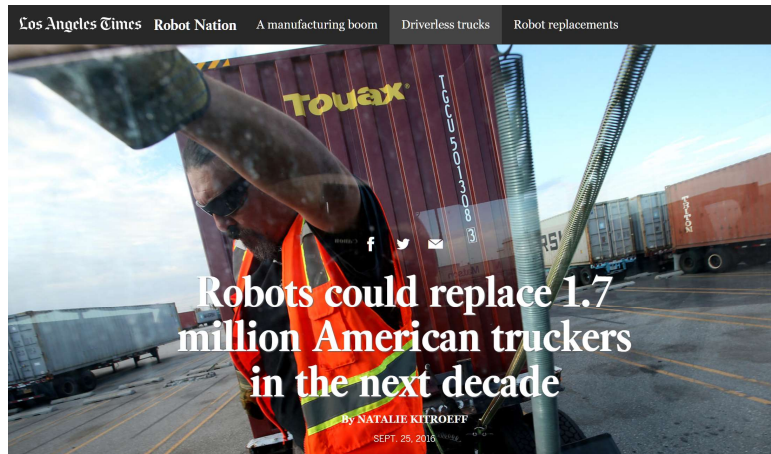
Ethical Issues

- Examples of ethical issues
 - Making good decisions (value alignment)
 - Honesty
 - Safety
 - Fairness
 - Having a positive social and economic impact

22

22

Jobs



23

23

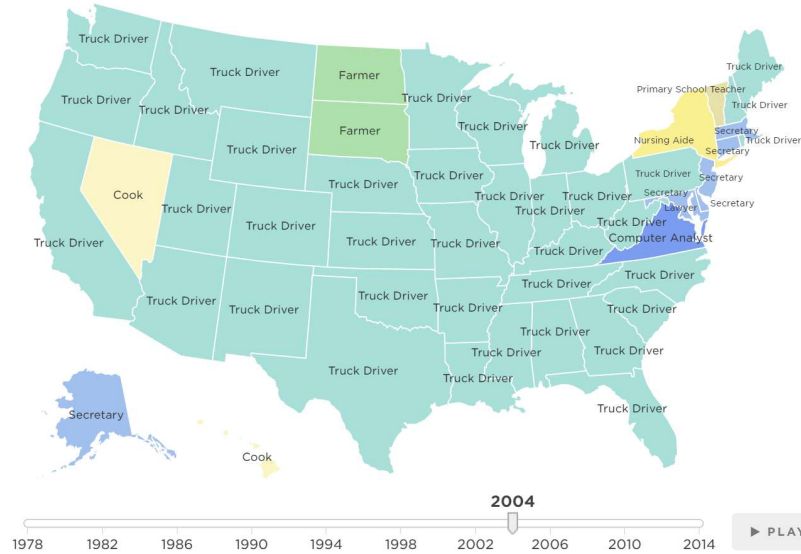
Jobs

- AI systems raise ethical concerns since ...
 - AI techniques are increasingly commercialized.
 - AI techniques are often used for automation.
 - AI techniques can result in cheaper, better performing, more adaptive, more flexible, and more general automation solutions than more traditional automation techniques.

24

24

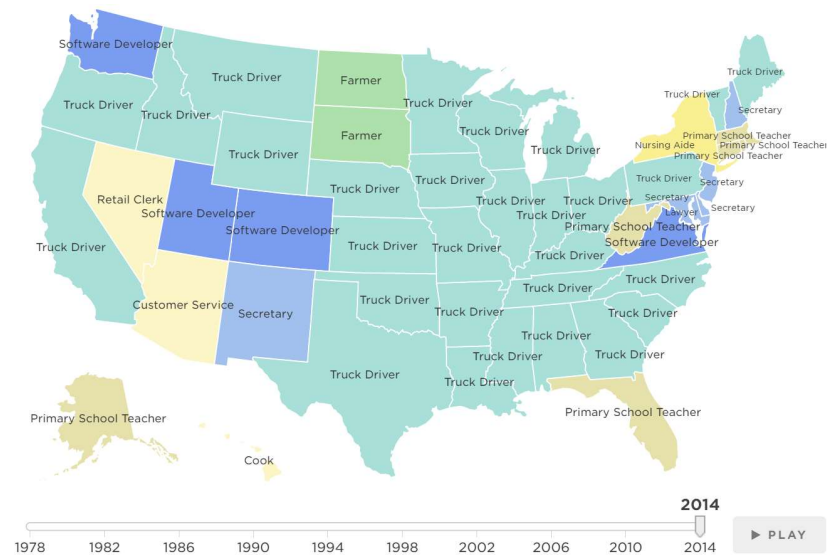
Jobs



25

25

Jobs



26

26

Jobs



The women of the Jet Propulsion Laboratory helped launch the first American satellites, lunar missions and planetary explorations. Those "human computers," as they were called, are seen here in 1953.



<https://www.npr.org/2016/04/05/473099967/meet-the-rocket-girls-the-women-who-charted-the-course-to-space> (Courtesy NASA/JPL-Caltech)

27
<https://wowpencils.com/ti-84-plus-ce-review/>

27

Summary

- Examples of ethical issues
 - Making good decisions (value alignment)
 - Honesty
 - Safety
 - Fairness
 - Having a positive social and economic impact
 - Jobs

28

28