Introduction to AI Ethics

- with a design twist

Aalto University - Information Technology Program
Juho Vaiste, 16.8.2018

About myself

- Academic:
  - M.Sc. in Business, “drop-out” of Philosophy studies
  - Multidisciplinary-lover, theoretical and philosophical approach
  - Work group by the Finnish Gov, Turku AI Society chair, Millennial Board AI member
- Entrepreneur
  - Before AI ethics: 10 years of entrepreneurship, 7 years in the IT industry
  - Consulting AI, technology and data ethics
  - Working in a project forecasting traffic flows and air quality data

TURKU AI SOCIETY

What is AI?

- By artificial intelligence, it is usually referred to an agent or a system which can operate tasks and functions to be thought earlier to be only in the reach of human intelligence.

- Machine learning is a popular and quickly developing subfield of AI industry, and it is already changing our habits in problem-solving and decision making.

TURKU AI SOCIETY

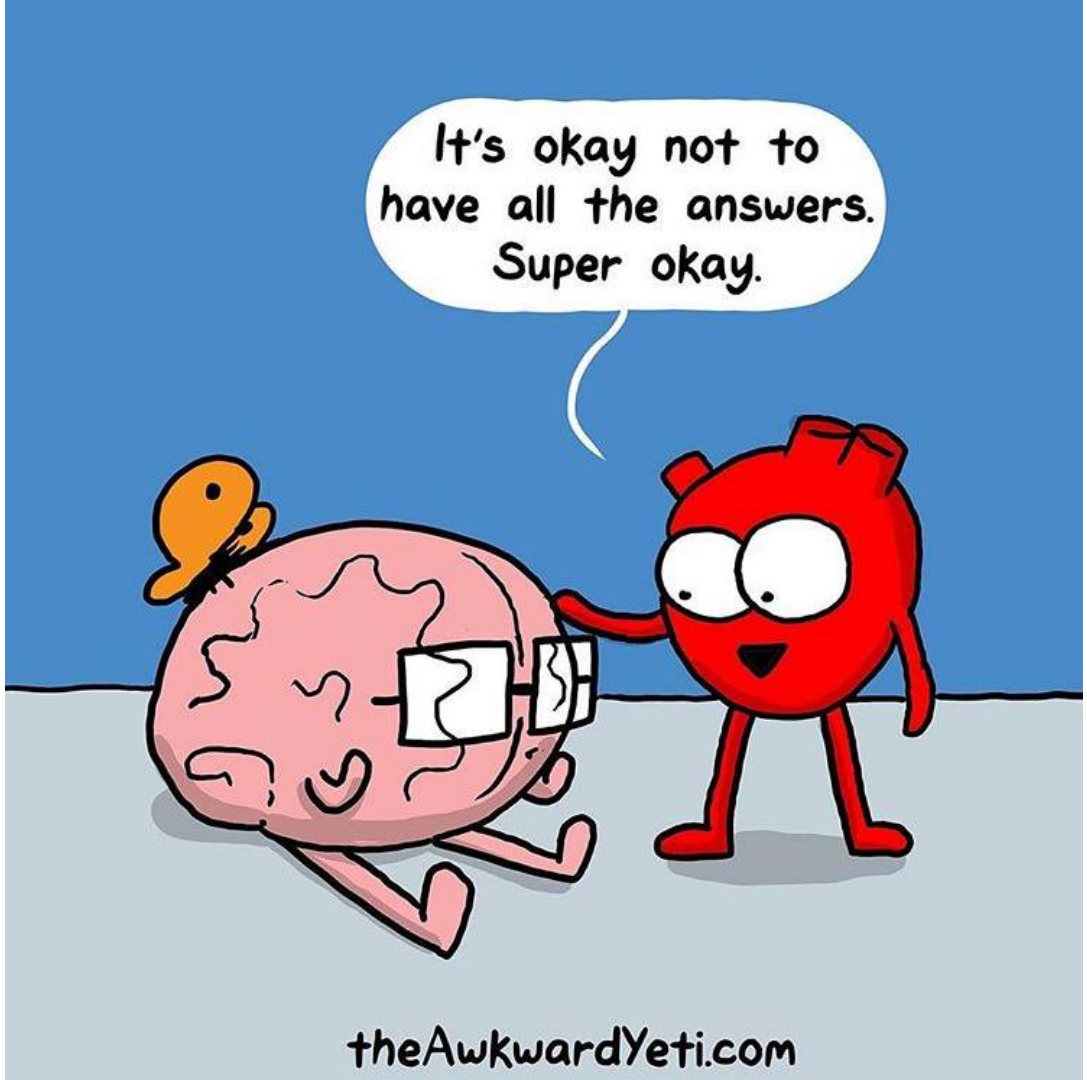
What is ethics?

- *"The field of ethics (or moral philosophy) involves systematizing, defending, moreover, recommending concepts of right and wrong behavior."*

- Isn’t that easy?

Ref: The Internet Encyclopedia of Philosophy - James Fieser, Ethics. https://www.iep.utm.edu/ethics/

TURKU AI SOCIETY


It's okay not to have all the answers. Super okay.
theAwkwardYeti.com

AI Ethics?

"What does it mean for an AI system to make a decision? What are the moral, societal and legal consequences of their actions and decisions? Can an AI system be held accountable for its actions? ..." Dignum, V. Ethics Inf Technol (2018)

- An emergent field of analyzing, researching and promoting the ethical consequences, concerns and problems related to AI technologies

TURKU AI SOCIETY

Ref: Dignum, V. Ethics Inf Technol (2018) 20: 1. https://doi.org/10.1007/s10676-018-9450-z

1. Warm-up workshop

Current challenges related to AI technologies

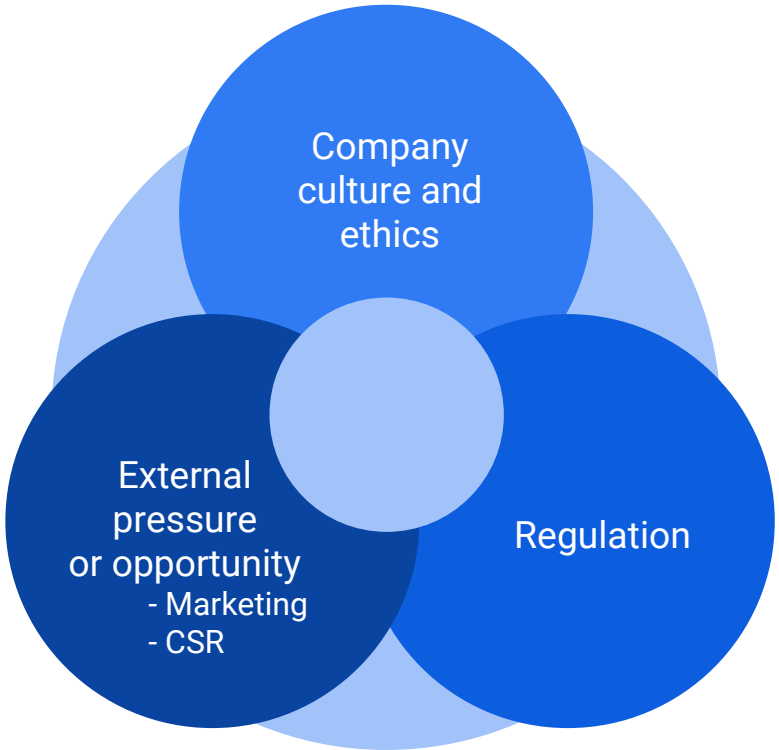
Mapping the background:
Ethics and the role of companies

- Ethics: philosophy, practical ethics (technology ethics), business ethics
- Technology philosophy
- Law, regulation, norms

- Corporate social responsibility
- Sustainability and the climate: starting 50s, highlighted during 21th century
- Technological responsibility is like to be the next “trend”
- In economics: externality costs with possible catastrophic consequences

TURKU AI SOCIETY

Mapping the background: Why should companies be interested in AI Ethics



TURKU AI SOCIETY

Mapping the background: technology

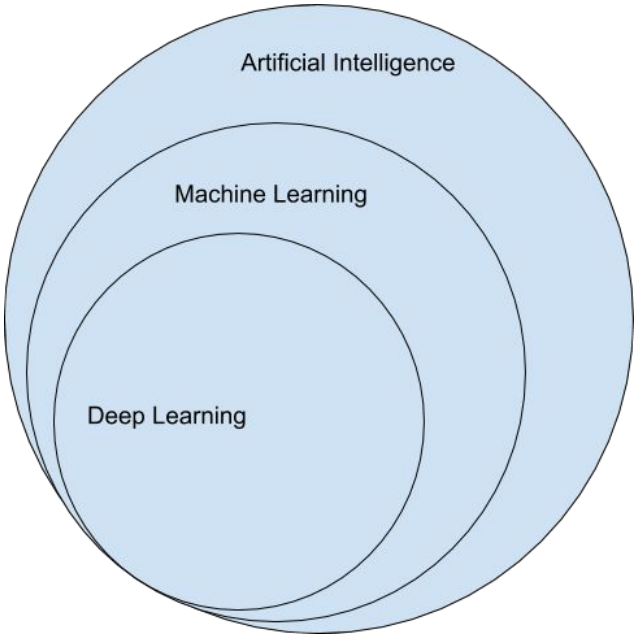
- Technology: internet, social media, data, devices, IoT

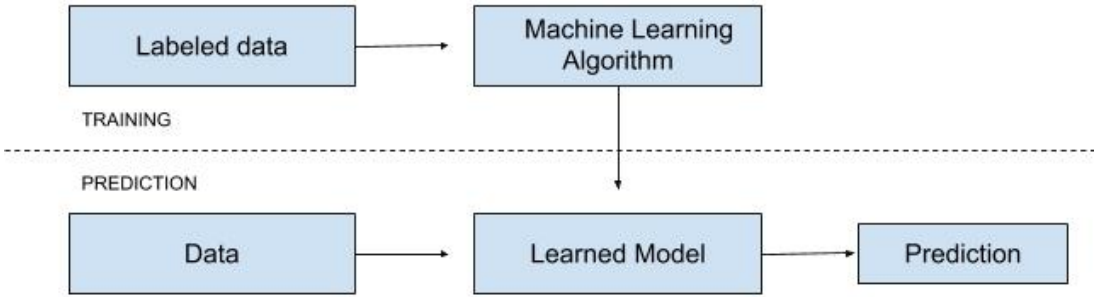
- AI: Definition under debate. Autonomous nature as a key driver.
- Something new, a lot of old: AI didn't just appear from nothing, but is based on the longer digitalization and technological progress

- Current (narrow) AI: Machine learning, deep learning, data-powered, narrow application areas

TURKU AI SOCIETY

Mapping the background: technology





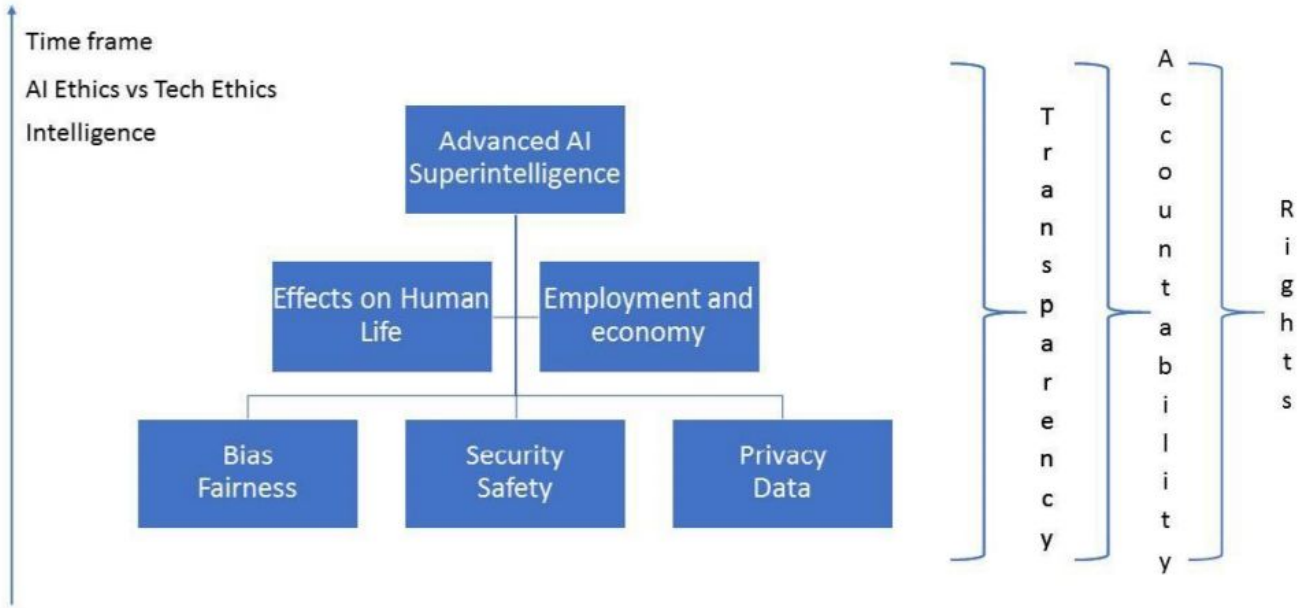
Mapping the background: Current AI

Application areas to keep eye on

Self-driving vehicles: concrete discussion on the responsibility issue
Weapons, war industry: most critical area of safety and security
IT industry: the development of AI and using and building the data
Healthcare: responsibility issues, but clear positive impacts
Professional services: hitting the employment issue the hardest
Gaming and entertainment: issues with addictivity

TURKU AI SOCIETY

Conceptual framework for ethical risks and concerns related to AI technologies


Time frame
AI Ethics vs Tech Ethics
Intelligence
Advanced AI Superintelligence
Effects on Human Life
Employment and economy
Bias Fairness
Security Safety
Privacy Data
Transparency
Accountability
Rights

Ethical risks (issues and concerns)

Data centralization and privacy

- *Established approaches to privacy have become less and less effective because they are focused on previous metaphors of computing, ones where adversaries were primarily human.* (AI Now, 2017).

- **Not only fancy new:** many ethical concerns are present in earlier technologies, but strengthen with AI
- **Data centralizing to the hands of few**
- **Surveillance, rating humans**
- **Using your data:**
    1) you should be able to decide
    2) which are the general limits

TURKU AI SOCIETY

Security and Safety

Core concepts: Unintended harm and malicious actors

- **Concrete problems:** questions of the technical development, but context-based understanding and design is needed
- **Change in the mindset:** understanding technological safety and security
- **Problem of responsibility**: a major topic and to be discussed in every application field

- **RL areas:** military robots, autonomous cars, IoT-devices
- Privacy vs security: do we need to give up our privacy for secure society?

TURKU AI SOCIETY

Bias and Fairness

- Algorithm and data biases can cause discrimination and unfair decisions.
- Algorithms and data can easily contain biases whose origin is in human decision-making, and these biases are difficult to detect, perceive and fix.

- **Sources for bias:** data used to train the machine learning model, human mistake (cultural assumptions), lacking diversity in the development or design team
- **Problems with data:** incomplete, biased or skewed, drawing on poorly defined and non-representative samples

TURKU AI SOCIETY

Effects on human life

“The technologies and the systems they enable are rapidly shifting behaviours and creating new rules for human interaction by virtue of incentives and boundaries built into their design.” (World Economic Forum, 2018)

- Critic towards digitalized life
- What is valuable, how do we want to build our lives
- The growth of mental disorders
- Building and finding meaning for ourselves
- Sex robots and relationships
- Cost diseases
- Surveillance, efficiency of human life

TURKU AI SOCIETY

Economic impacts and employment

- The change and challenge AI sets for our employment and distribution of income is potentially enormous

- **Employment:** you don't need to be a luddite to believe in radical change
- **Distributing income in a new ways:** active research and experimenting on basic income is needed
- **Meaning:** We have used to built a great part of our meaning throughout our work.

TURKU AI SOCIETY

Lenses and principles to approach the risks

Transparency

- Transparency means that the stakeholders - the public, users, developers, owners - understand how machine learning based systems work and what is their decision-making process including

- Emphasized especially by the public sector: crucial for public decision-making that all the decisions made by machines are transparent, predictable and fair, manipulation-free and the responsibility sharing is planned well.

TURKU AI SOCIETY

Accountability & responsibility

- *“Accountability in this context includes an obligation to report, explain, or justify algorithmic decision-making as well as mitigate any negative social impacts or potential harms.”*

- *“Who is responsible if users are harmed by this product?*
  What will the reporting process and process for recourse be?
  Who will have the power to decide on necessary changes to the algorithmic system during design stage, pre-launch, and post-launch?”

- Reference/See: Principles for Accountable Algorithms and a Social Impact Statement for Algorithms http://www.fatml.org/resources/principles-for-accountable-algorithms

TURKU AI SOCIETY

Rights

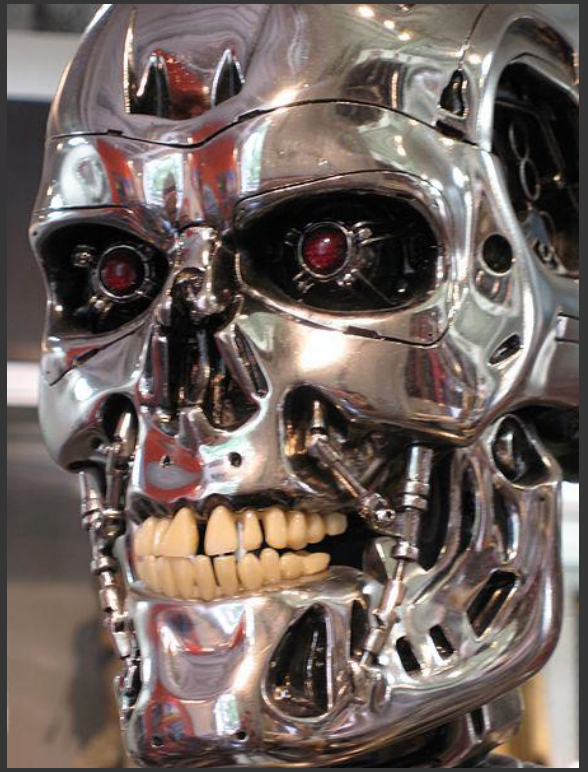
- Rights are essential to ethical analysis and thinking
- Concrete fundamentals for our societies

- AI technologies should not contravene or violate our common rights
- UN Human rights works as a fundamental starting points, but which other rights we should respect and in which situations?

- From the perspective of company, cultural and area-specific knowledge is critical

TURKU AI SOCIETY

… that's AI ethics before superintelligence



Advanced AI systems and superintelligence

- Philosophically interesting topics
  - Human-level intelligence, superintelligence
  - Robot rights
  - Human agency

- In practice (have to be taken seriously, but far away from practice)
  - Not a question in the current AI development
  - Academically AGI related responsibility of corporations might become a topic in next years
  - To relieve the worries: there's people working on with AGI safety and alignment

TURKU AI SOCIETY

Sources: institutions, universities

- AINow (New York University)
- AI100 (Stanford)
- The Royal Society
- World Economic Forum
- Internet Society
- MIRI (Machine Intelligence Research Institute)
- Future of Humanity Institute (Oxford)

- ACM (Association for Computing Machinery)
- Tivia (Finland)
- IEEE

TURKU AI SOCIETY

Finnish Community of AI Ethics

- Academic
  - Helsinki: MOIM, Rajapinta, HY, Aalto
  - Turku: Turku AI Society, Future Ethics group
  - Tampere: Rose project, Laitinen & Parviainen et al
  - Jyväskylä: IT Ethics group, Vakkuri

- Government: strategy work groups, Tekoälyaika project
- Organizations: TIVIA, Millennial Board AI

- National Seminar of Theoretical AI

TURKU AI SOCIETY

AI Ethics in Practice

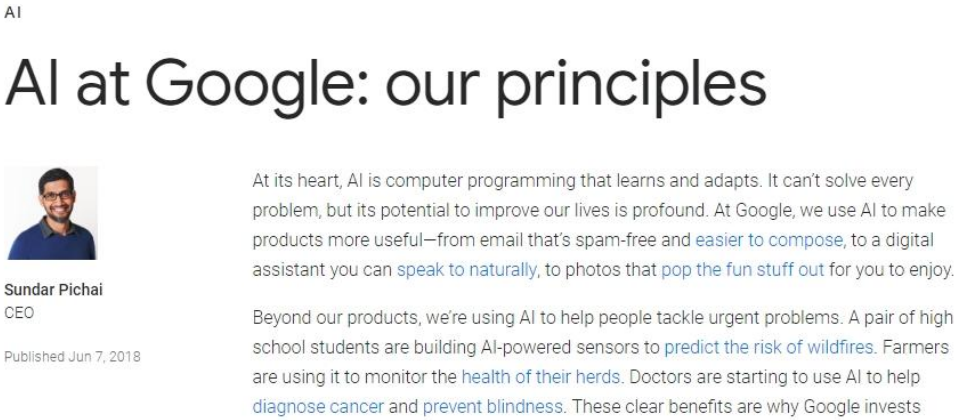
Design, tools & frameworks

AI Ethics in Practice

Right now:

A lot of work is done at the strategy level and companies are creating ethical principles for their AI technology

The first step is to define the strategic guidelines



AI

AI at Google: our principles

Sundar Pichai
CEO

Published Jun 7, 2018

At its heart, AI is computer programming that learns and adapts. It can't solve every problem, but its potential to improve our lives is profound. At Google, we use AI to make products more useful—from email that's spam-free and easier to compose, to a digital assistant you can speak to naturally, to photos that pop the fun stuff out for you to enjoy.

Beyond our products, we're using AI to help people tackle urgent problems. A pair of high school students are building AI-powered sensors to predict the risk of wildfires. Farmers are using it to monitor the health of their herds. Doctors are starting to use AI to help diagnose cancer and prevent blindness. These clear benefits are why Google invests

Ref: https://www.blog.google/technology/ai/ai-principles/

TURKU AI SOCIETY

AI Ethics in Practice: Principles (Fin AI Company: Fourkind)

#1: To the extent that it is possible, avoid creating AI systems that reinforce commonly understood social biases, contravene human rights, violate human dignity, or otherwise breach your own moral code

If you encounter and/or are asked to do something you find unethical, raise the issue immediately internally and with your client. Never sacrifice your own values for added model accuracy, financial gain, or fear of failure.

#2 Follow all applicable laws, directives and mutually agreed-upon best practices

This also applies to things like internal and client-company privacy guidelines/policies and GDPR. If you are unsure about interpretation, raise the issue with your client or internally, depending on the circumstance. Note that in some cases, such as laws regarding minors, you may have to *enforce* discrimination to behave ethically.

#3: Hold yourself accountable

If you develop an AI system, take ownership of it. Be honest about the way it works, the type of data used to train it, and address issues in a timely fashion. Don't place the blame on others or try to hide mistakes. Inform clients immediately if something goes wrong.

#4 Be as transparent as you can

The inner workings of some AI models can be explained easily, some are more of a black box. Sometimes laws and regulations affect the choice of learning algorithm. In all cases, make sure to document the data and process you use when developing AI. Transparency also applies to end-users; make sure that the use of AI is documented and to offer a non-AI alternative solution not only where required, but also where feasible.

#5 Use your own judgement

At Fourkind, we go to great lengths to find the best people we can; people who we trust to make sound decisions not only collectively, but also as individuals. Don't be afraid to speak up and err on the side of caution when you feel something isn't quite right.

Ref:
https://medium.com/value-stream-design/fourkinds-thoughts-on-the-ethical-use-of-artificial-intelligence-85672585d3f1

AI Ethics in Practice

- From philosophy and strategy to practice
- Design - Build - Evaluate

- Ethical design as a key function
- *“Ethical design means considering the context of the product you create.”*


FUTURE TENSE

What Are “Ethics in Design”?

People in the field are calling for more ethical decision-making—but it’s hard to pin down what exactly that means.

Reference: https://slate.com/technology/2018/08/ethics-in-design-what-exactly-does-that-mean.html

TURKU AI SOCIETY

AI Ethics in Practice

- Ethics and AI should be linked at many levels:
    - Ethics by design: ethical reasoning as a part of the behaviour of AI agent
    - Ethics in design: methods and processes supporting ethical evaluation of the societal and structural changes AI systems has on our societies
    - Ethics for design: standards and guidelines to ensure the ethical development of AI

TURKU AI SOCIETY

Reference: Dignum, V. Ethics Inf Technol (2018) 20: 1. https://doi.org/10.1007/s10676-018-9450-z

Tools and Frameworks for Ethical Design

- Traditionally in applied ethics: virtue ethics, duty ethics and consequentialist ethics
- ...or normative principles

- AINow: Algorithmic Impact Assessments (for public algorithms)
- The Ethical Matrix (Oneilrisk.com, Cathy O'Neil)
- Iyad Rahwan: "Society-in-the-Loop: Programming the Algorithmic Social Contract"
- Ethics Kit - Design Better (ethicskit.org)

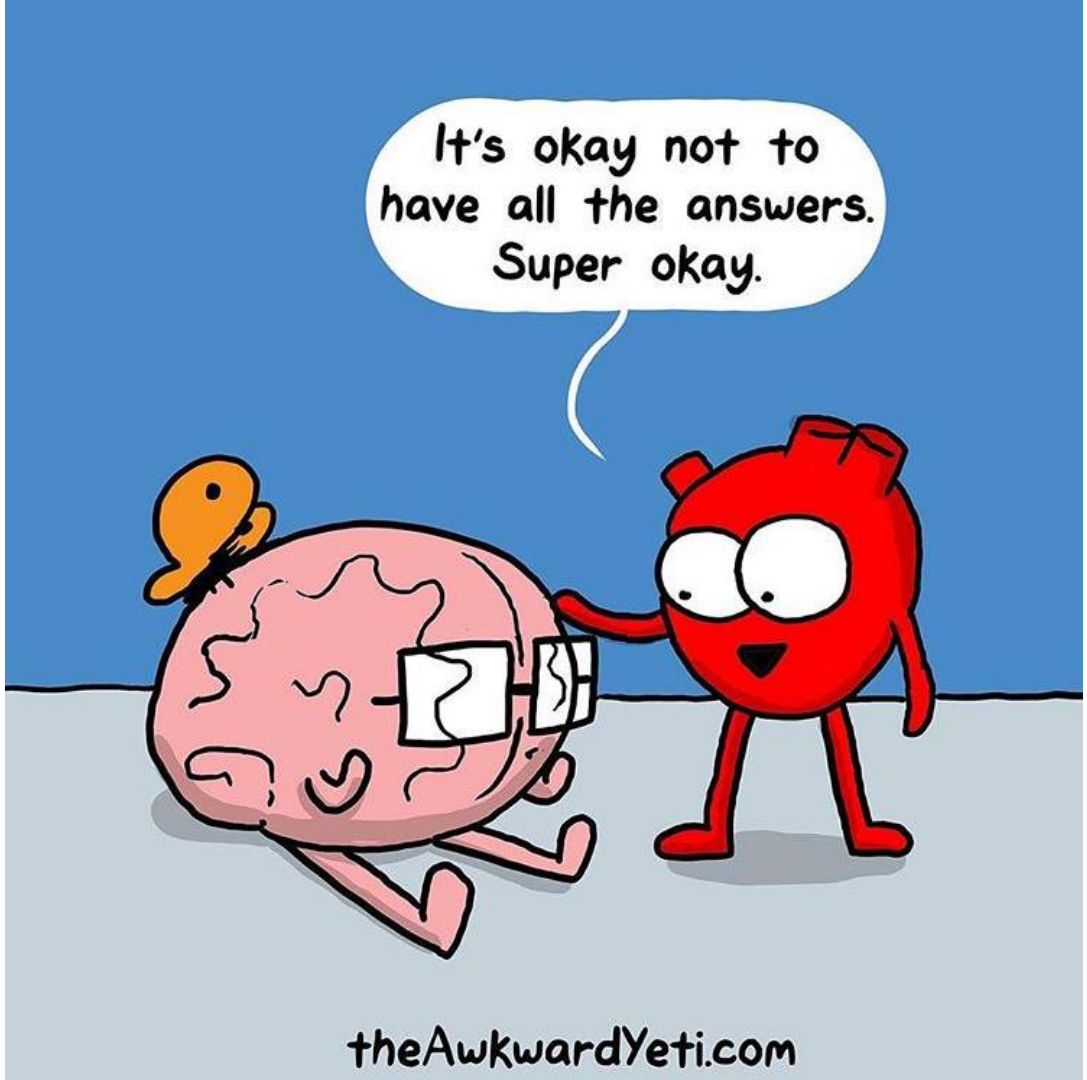
TURKU AI SOCIETY

Two things before the workshop

1. **There is not right answers to all questions...**
   - Hear everyone and give everyone a chance to speak
   - Respect each others' opinions and be polite
   - Embrace ethical thinking
   - Even though your group would agree with the issue, consider other possible opinions and prepare to accept that your view might not be universal

2. **... but there is a great deal of material and tradition to built on (especially in real-life situations)**
   - Understand and analyze what has been said earlier
   - Reports, guidelines and standards are a great starting point for AI ethics work

TURKU AI SOCIETY


It's okay not to have all the answers. Super okay.
theAwkwardYeti.com

2. Workshop

Design tools for ethical and responsible AI

2. Workshop

1. Take your groups (max 10 persons)
2. Choose your case. Take another than in the workshop 1
3. Choose your group's a) chairperson b) presenter
4. The idea is to go through the Ethics Cards together and choose cards which feel relevant to your case (5-8)
5. Discuss together each of the selected cards, and use notes to record your considerations and conclusions
6. Build a logical cluster from your cards and notes, and present!

TURKU AI SOCIETY

Presenting the results of the 2nd workshop

What's next?

- **Spread the awareness.** AI ethics includes questions we have to solve as societies.
- **If you got interested in the topic professionally/academically,** there is a good listing of learning resources for continuing your learning. AI ethics courses are popping up at universities around the globe.
- **Questions?** Feel free to contact in any AI ethics matter. LinkedIn & Twitter: juhovaiste

- **Thank you and enjoy your time in Helsinki!**

TURKU AI SOCIETY

Extra readings and study materials

Great and free introduction course to AI (by University of Helsinki: http://www.elementsofai.com/fi

Top 9 ethical issues in artificial intelligence
World Economic Forum
https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/
The Future of Employment: How Susceptible Are Jobs to Computerisation?
Carl Benedikt Frey, Michael A. Osborne
https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf
Concrete problems in AI safety
Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, Dan Mané
https://arxiv.org/abs/1606.06565
The Ethics of Artificial Intelligence
Nick Bostrom, Eliezer Yudkowsky
https://nickbostrom.com/ethics/artificial-intelligence.pdf
Research Priorities for Robust and Beneficial Artificial Intelligence
Stuart Russell, Daniel Dewey, Max Tegmark
https://futureoflife.org/data/documents/research_priorities.pdf
Fairness, Accountability, and Transparency in Machine Learning
https://www.fatml.org/

TURKU AI SOCIETY

Extra readings and study materials

Stanford - One Hundred Year Study on Artificial Intelligence
2016 report
https://ai100.stanford.edu/2016-report
Artificial Intelligence: A Modern Approach (3rd Edition)
Stuart Russell, Peter Norvig
https://www.amazon.com/Artificial-Intelligence-Modern-Approach-3rd/dp/0136042597
The Cambridge Handbook of Artificial Intelligence
http://www.cambridge.org/gb/academic/subjects/philosophy/philosophy-mind-and-language/cambridge-handbook-artificial-intelligence?format=PB#uCBu0mtFhibOGaY6.97
Nature: Anticipating artificial intelligence
Why we need AI research beyond the technological research
https://www.nature.com/news/anticipating-artificial-intelligence-1.19825
Rise of the Robots: Technology and the Threat of a Jobless Future
Martin Ford
https://www.amazon.com/Rise-Robots-Technology-Threat-Jobless/dp/0465097537

O'Neil C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy.

Machine learning: the power and promise of computers that learn by example (2017). The Royal Society.

TURKU AI SOCIETY

Thank you!

aisociety.fi
juhovaiste.fi
@juhovaiste

References

The ART of AI — Accountability, Responsibility, Transparency. Virginia Dignum.
https://medium.com/@virginiadignum/the-art-of-ai-accountability-responsibility-transparency-48666ec92ea5

Dignum, V. Ethics Inf Technol (2018) 20: 1. https://doi.org/10.1007/s10676-018-9450-z
https://link.springer.com/article/10.1007/s10676-018-9450-z

The Internet Encyclopedia of Philosophy - Ethics. https://www.iep.utm.edu/ethics/

AI Now (2017) https://ainowinstitute.org/AI_Now_2017_Report.pdf

What Are “Ethics in Design”? Slate Magazine. Victoria Sgarro, 13082018.
https://slate.com/technology/2018/08/ethics-in-design-what-exactly-does-that-mean.html

AI at Google: our principles. Sundar Pichai. 07062018 https://www.blog.google/technology/ai/ai-principles/

Fourkind’s thoughts on the ethical use of artificial intelligence. Max Pagels, 10082018.
https://medium.com/value-stream-design/fourkinds-thoughts-on-the-ethical-use-of-artificial-intelligence-85672585d3f1

References

Agile Governance Reimagining Policy-making in the Fourth Industrial Revolution. World Economic Forum (2018). http://www3.weforum.org/docs/WEF_Agile_Governance_Reimagining_Policy-making_4IR_report.pdf

Ethics Kit http://ethicskit.org/ethics-cards.html

The Awkward Yeti comic http://theawkwardyeti.com/ *"Publications: If your publication is completely not for profit, you may use my web-based comics any time, as long as my URL is still the primary attribution on the image."*

Terminator photo https://www.flickr.com/photos/tenaciousme/540330647/ https://creativecommons.org/licenses/by/2.0/

Principles for Accountable Algorithms and a Social Impact Statement for Algorithms http://www.fatml.org/resources/principles-for-accountable-algorithms