

AI and Ethics

Richard Wheeler
Reykjavik University



My Role

I'm one of four instructors in this course because *AI and Ethics* covers a lot of ground, from philosophy and logic to legal and regulatory frameworks. This is the first time the course will be taught, and so we are still working out the best possible methods for teaching and curriculum.

Ethical and responsible design and use of AI systems is complex and participatory; in your career you will often have to make decisions about ethics with other people, and in a group setting, so we have tried to include that in the course design.



My Role

This course is meant for you to learn and engage, not memorise facts and methods. We want you to understand the issues involved, methods in AI and ML that might have negative impacts, and know how ethics are understood in real life. In this context, discussing use cases is a primary method of instruction. As in ethics itself, there are sometimes no right answers, only a walk through the right questions.

My role of the four instructors is primarily to communicate the realities of ethics in modern research, science, and business as related to AI. After me, you will learn more about the philosophy of ethics, and about automated reasoning systems in the context of on-going work within the AI group at RU.



My Role

Each class will end with a short *ask me anything* session. Feel free to ask anything on any subject. It seems like everyone here has a different background: some of you are coding geeks, others love philosophy or policy and economics, or civil services. All are valid paths, and all need to know why AI is an ethical minefield.

In this context, there are no stupid questions, so ask anything. I will try to answer as honestly and directly as I can, though sometimes I will have no idea what a good answer is. :>)



Some Truth.

The truth is, or the reality is, the planet is in serious trouble. We have destroyed the environmental systems that have sustained life on Earth for 3.7 *billion years*.

We have disrupted and irreversibly changed systems so complex no human could ever likely comprehend them. In a real sense, AI might be the planet's only long-term hope for saving humanity and life on the planet. That's a lot of responsibility.

Our job now is to design, *ethically*, the systems our children and grand-children will need to save the planet, and many of these systems and tools will be based on AI.

This is not about what books Amazon recommends, or what kinds of images an AI system generates for entertainment. This is about the race between science and catastrophe, with the future of life on Earth hanging in the balance.



AI and Ethics

In the context of artificial intelligence, ethics has a few main concepts to understand:

- What ethics means, why it is important in the systems we build and use.
- That humans are not always very ethical in their behaviour, and we likely want our creations to be smarter, and more ethical, than ourselves. (or do we?)
- That what we create may find unintended uses that make the ethical, unethical.
- That we may create unethical systems without realising it.
- That ethics in complex AI systems is partially an engineering problem: you need to understand information theory, search and boundary condition concepts, robust design, etc.
- That AI systems can be used unethically, or act unethically, and this is a real problem.
- That there are fundamental philosophical roots to ethics, but in your professional career you will have to make ethical decisions where there is not likely a best solution known.
- That acting ethically in your professional life and with AI requires you to think and act *wilfully*.



AI and Ethics

Generally, we want AI to:

- Never cause harm
- Be accountable and transparent
- Be robust to new situations and unseen circumstances and uses
- Protect and encourage fundamental human rights
- Act in accordance with the best practices and ethical and legal guidelines of the domain in which it was created
- Be able to explain its reasoning, and to accurately account for certainty and confidence in its data, knowledge, and decisions
- Know when it cannot make a decision, and refuse to make unethical decisions even when asked



Levels of Ethics

Individual ethics: morals, culture, religion, education



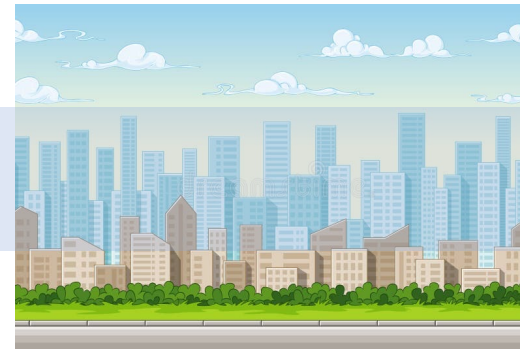
Community of practice: guidelines, best practices



Institutional ethics: organisation, legal, missions, compliance

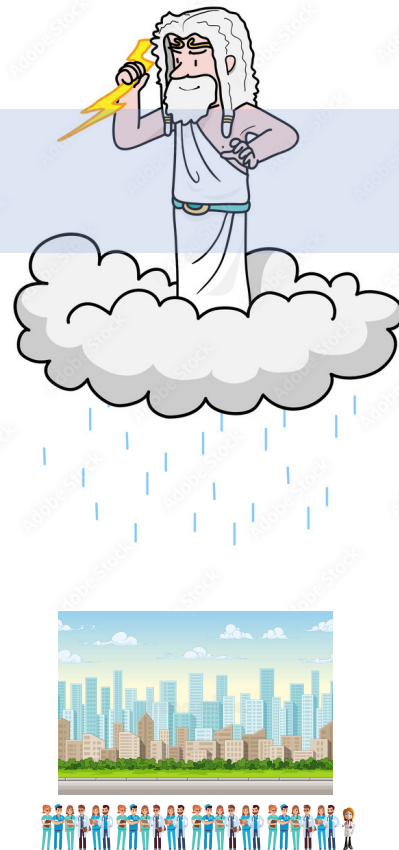


Societal ethics: traditions, laws, regulations, policies, governing bodies



Levels of Ethics

Philosophical ethics: what does it mean to be ethical?



Some Differences...

Usually, *ethical* means *compliant with social expectations for good behaviour in relation to the people around you*. Sometimes society codifies these expectations into their cultures, sometimes into their laws, and sometimes into their religions.

Reality usually means the universe that exists without human perception – it is whatever are the real laws that govern our cosmic instance whether we know them or not.

Truth is not reality; truth is consensual – it is what the group around you thinks it is. People do not make ethical decisions based on reality, but on how their society and they themselves perceive *truth* and their circumstances. Truth, like ethics, is mostly situational, it changes with circumstances, culture, and era. Or does it?



Professionally...

An important distinction is that you will not be asked to act *morally* in your careers (though I would ask this), you will be asked to be ethically *compliant* with rules, guidelines, and legislation, which is in some ways a different thing. It is the difference between *why* we need to act ethically, and *how* we act ethically in relation to our obligations to society.

Professionally, in the real world ethical behaviour means:

- Knowing and abiding by the legal requirements and guidelines in your field
- Avoiding blame and responsibility for your actions, and not causing your employer any trouble
- Acting in a way that can be defended and explained, whether truthful or not

But we want to be better than that, don't we? The difference between generations, the arc of history. Your generation cannot afford any more unethical conduct...



A Simple Example

I helped an academic colleague with their European Research Council proposal, and they had more than 700 publications in physics in the last seven years.

Doing the math, you can see there is a problem here. They published one paper every three days or so, and they were very proud of this.

This colleague is part of a nuclear research consortium in which all work and publications are shared. He was very proud of the 700 publications, and was genuinely shocked when I asked him if he thought it was ethical to claim all of these in his H index and CV. He thought it was fine.

What do you think – ethical, unethical? Neither?



The Result...

So what happened?

I asked this colleague how many of these publications he had *read*, not *wrote*, *read*. He said maybe 20 or 30 out of 700. Put in this context, is it ethical to claim professional credit for 670 publications they have not even *read*? Now it is sounding a lot more unethical. So we had to engineer a way to cast these publications as a good, not a negative, thing...

It is now common for publications to be written by AI, faked, or with fake data and results, or with co-authors who have not even read the paper as a way of gaming academia.

Ethical?



This Year in MSCA PF

Guidance on the use of generative AI tools for the preparation of the proposal

When considering the use of generative artificial intelligence (AI) tools for the preparation of the proposal, it is imperative to exercise caution and careful consideration. The AI-generated content should be thoroughly reviewed and validated by the applicants to ensure its appropriateness and accuracy, as well as its compliance with intellectual property regulations. Applicants are fully responsible for the content of the proposal (even those parts produced by the AI tool) and must be transparent in disclosing which AI tools were used and how they were utilized.

Specifically, applicants are required to:

- Verify the accuracy, validity, and appropriateness of the content and any citations generated by the AI tool and correct any errors or inconsistencies.
- Provide a list of sources used to generate content and citations, including those generated by the AI tool. Double-check citations to ensure they are accurate and properly referenced.
- Be conscious of the potential for plagiarism where the AI tool may have reproduced substantial text from other sources. Check the original sources to be sure you are not plagiarizing someone else's work.
- Acknowledge the limitations of the AI tool in the proposal preparation, including the potential for bias, errors, and gaps in knowledge.



But First.

Something no one will ever say.

Like every other part of biological existence and perception, your inner morals and orientation are dominantly genetically determined, though your *behaviour* may be modified by social and legal pressures and expectations. Morals and ethics are not exactly the same thing. People born for whatever reason without morals can still act ethically, but does that make them an ethical person?

For example, a brief test. If you have heard this one, do not give away the ending. :>)

The psychopath test...



Even Stranger.

Something else no one will ever say.

Probably the majority of history's prominent figures were extremely mentally ill, and the mentally ill dominate politics and the business world.

When sociopaths are in charge and running everything, what do ethics even mean?

How bad people – those acting in bad faith, win.



Definition.

One of Google's first mottos was "don't be evil". But is there even a practical definition of evil, that might be useful when considering ethics?

H.G. Wells, around 1940: *"Evil is the use of power against the powerless."*

In some ways this is at the heart of many AI and ethics problems. The technology is powerful and the people are not. The people have become transparent, but the technology and processes are not.



Definition.

This is an important point. Much of modern ICT, including AI and ML, the internet, the web, social media, are all built on extreme inequalities of power, information, and wealth. For most of the companies in this sector, you are not their *customer*, you are their *product*. ChatGPT and LLMs were not built for you, they were built to replace you.

The point is that most internet-related technologies do not make the world better, but make it more unequal. Human needs are not important to the world of internet companies other than as a possible product to market. This makes most instances of new technologies driven forward by uncivil motives and unethical means. Example: *those scooters everyone loves – you know, like HOPP.*

For example, I can order food on my phone from the smallest café in India (Smally's Restocafe), but that same phone cannot tell me how to get out of the building we are in if there is a fire. It cannot tell me how to reach this room if I am in a wheelchair. It cannot tell the rescue services how many people were in the building when it collapsed in an earthquake. Those services do not exist because...?



Definition.

So if you want to know if a system is *unethical* (as opposed to *ethically compliant*) look for extreme asymmetries in wealth, power, and information. This goes for AI systems as well.

Consider most insurance companies. They know every detail of your finances, lifestyle, sexuality, your habits, even when you are likely to die. But when you call them on the phone, the person at the insurance company typically will not even give you their real name. That's *asymmetry*. They know everything about you, and you are forbidden from knowing anything about them.



Question?

(Those big companies you think are your friends) your bank, etc., all usually know when you are likely to die well before you do.

Examples from real life, purchasing habits, data mining, language-independent speech analysis in SAAM. A friend in Italy, a family member in Indiana...

Do these companies have any obligation to tell you? What do they do in practice? Employers, insurance companies, hospitals...

And of course it is situational and cultural: in China, doctors tell an elderly person's relatives that they are dying, but the elderly person never knows. It's considered inhumane to tell the elderly person they are dying.



In the News...



The Verge / Tech / Reviews / Science / Entertainment / AI / More +

META / TECH / GOOGLE

Meta and Google secretly targeted minors on YouTube with Instagram ads



Cath Virginia / The Verge | Photos from Getty Images

/ The companies reportedly made use of a loophole to intentionally serve ads to kids.

By [Jess Weatherbed](#), a news writer focused on creative industries, computing, and internet culture. Jess started her career at TechRadar, covering news and hardware reviews.

Aug 8, 2024, 10:33 AM GMT

[Link](#) [f](#) [t](#) | 26 Comments (26 New)




Speaking of Google

The Verge / Tech / Reviews / Science / Entertainment / AI / More +

META / TECH / GOOGLE

Meta and Google secretly targeted minors on YouTube with Instagram ads



/ The companies reportedly made use of a loophole to intentionally serve ads to kids.

By Jess Weatherbed, a news writer focused on creative industries, computing, and internet culture. Jess started her career at TechRadar, covering news and hardware reviews.
Aug 8, 2024, 10:33 AM GMT

Cath Virginia / The Verge | Photos from Getty Images

26 Comments (26 New)

Meta and Google teamed up to run a secret campaign that deliberately targeted 13 to 17-year-olds with Instagram ads on YouTube according to the *Financial Times*, breaking the search giant's own rules against advertising to children.

MOST POPULAR

Ethical?

Interestingly, research shows us that of the hundreds of people working in the tens of companies who were in on this programme, very few of them would have approved it as ethical themselves. If you asked them to approve it, they would not. And yet, together, they created a mass system for unethical conduct; and that's one of the main reasons why companies are primarily created, to do things people would not. People would not deny you medical care, but a company will.

Personal Ethics

Some stories about personal ethics, please give comments:

- When I realised what was really happening with climate change
- When my colleague realised what was happening with climate change
- My unfortunate nephew
- The limits of reasoning and ethics: my friend's departed mother – would he create heaven for her if he could?

Do Animals Have Ethics?



Yes, probably.

Do Animals Have Ethics?

The New York Times

Chimpanzees Show Skills in Managing Conflict

 Share full article



Certainly apes and monkeys seem to. Dogs as well.

Do Animals Have Ethics?

NewScientist

Sign in 

Enter search keywords

What do chimp 'temples' tell us about the evolution of religion?

By Rowan Hooper

📅 4 March 2016



They also have a form of religion.



Do Animals Have Ethics?



And complex social dynamics and behavioural codes.

Do Animals Have Ethics?



“Ha ha ha – you fell
down the stairs...”

But there might be
exceptions.

Do Animals Have Ethics?



But likely wherever there are groups of complex organisms, there is a kind of ethics.

In some ways, ethics is the behavioural code emerging from groups of sentient agents to keep the genome stable, cooperative, and competitive in the larger biome.

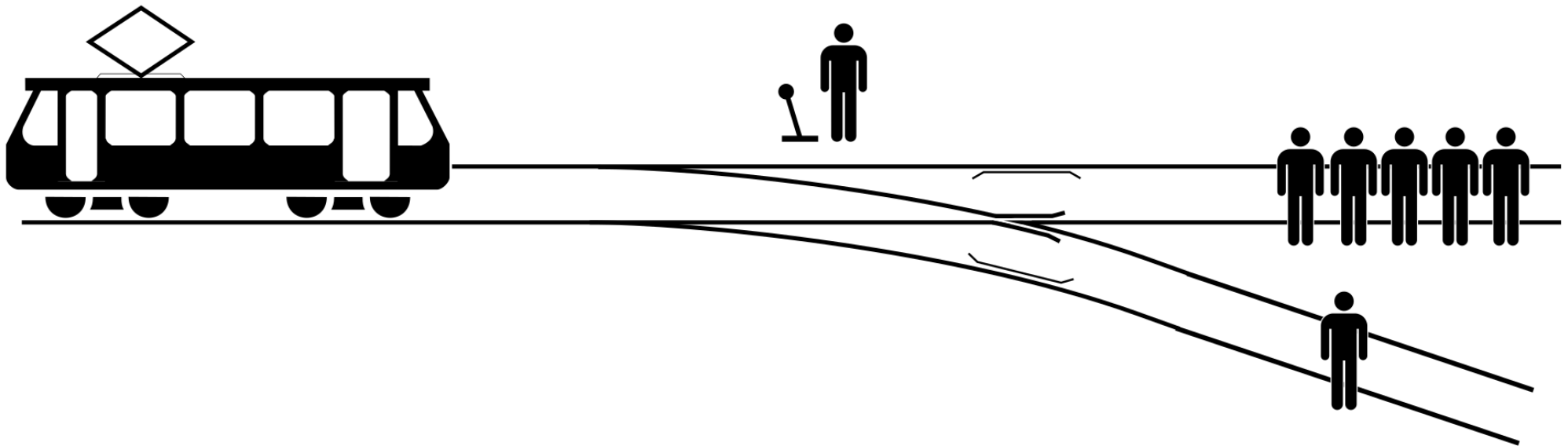
This means that ethics is a side-effect of evolution, and itself evolves to meet local environments and circumstances.

Ethics Evolve



*Ethics are not fixed, and evolve culturally, in an organisation, and personally.
Acting ethically is a personal journey, and the important thing is to learn and internalise good ethics and moral understanding.*

Responsibility.



Ye Olde Trolley Problem. This seems to be mostly about ethics, but it is also about personal responsibility and how society perceives us. Many of those who would not switch to the single person, would if they knew no one would ever know who operated the switch. I'd guess your answer to this question is determined at birth.

Psychology and Ethics

Your psychology directly impacts how you see ethics and ethical decision making.

An illustration.

People often fall into one of these categories:

- 1. See the world as composed of people*
- 2. See the world as composed of events*
- 3. See the world as composed of systems*



A Different Trolley Problem



What's the cause?

- 1. Drivers drunk?*
- 2. Freak accident?*
- 3. Systemic failure?*

A Different Trolley Problem



What's the cause?

1. *People* 80%
2. *Event* 15%
3. *System* 5%



All three.

Understand Your Weaknesses

People often fall into one of these categories:

- 1. Can't see the people*
- 2. Ignore the events*
- 3. See the world as composed of systems*

Personal Growth

For me:

- 1. Have had to learn to see the people that could make ethical decisions and act ethically or not*
- 2. Have had to learn to understand events that impact ethical decision making*
- 3. Acknowledge everything is not a systemic problem that can be engineered away*

Learn to tell the difference when problem solving and considering ethics.

Readings

How are you getting on with the readings? Are they too easy, hard, long, short, not the topics you are interested in? Would you prefer more scientific articles?

For Next Class

Ethical design of AI systems begins with problem identification and exploration of possible ethical issues. For August 29, come with a short idea for a mobile phone application that you can explain to the class very briefly. Nothing written required, and the ideas do not need to be good, just interesting. Each student will have about two minutes to present their idea, after which the rest of the class will give comments on design and possible ethical issues. The task is to identify and discuss ethical issues across a broad range of applied ideas.

For Monday September 2, write and submit around one page of analysis of the discussion: what idea you found most interesting, and discuss your own feeling for the ethical issues involved. The assignment is to show knowledge and insight into what constitutes ethical issues in AI systems. Explore what parts of the discussion you found most interesting given your background and scientific interests.

Ask Me Anything



Comments and discussion also on the online forum, of course.

Questions can also be sent to me directly at rwprivate@gmail.com.