AI in Education
KING'S COLLEGE LONDON

Ethics of generative AI

Dr Caitlin Bentley | Lecturer in AI Education

How do we create responsible and trustworthy AI technologies? What even is a responsible and trustworthy AI technology?

To answer these questions, we need to explore the complex ethical considerations of AI, and generative AI in particular. This topic is not only relevant to defining and designing responsible and trustworthy AI, but it's also important because it helps us to reflect on how AI impacts on our lives and society.

I want to point out that there's no one single solution to what makes AI ethical, or how it can benefit society more than it harms it. For example, recently, the Turing Institute in the UK asked people what they think about using facial recognition technology for things like quick airport security checks, or helping police. Seventy per cent of the people who were surveyed said it was beneficial, but about half were worried about job losses for border control, and the risk of police wrongly accusing innocent people. I think we also need to consider that when mistakes happen in policing, people who already tend to be discriminated against could be unfairly targeted even more.

This example highlights the nuances and complexities involved in ethical AI. One AI technology can lead to different concerns and harms for different people in different places, and in different applications of the same AI.

So my suggestion to you is to develop an awareness of the ethical issues with AI from a variety of perspectives. Be aware that there are multiple perspectives, with some that tend to get heard or recognised more or less than others.

I encourage you to avoid thinking there's only a right or a wrong way to approach things. It's crucial to understand the various viewpoints on AI ethics, including your own. Remember that everyone has a unique perspective shaped by different backgrounds. Trying to seek out and understand these different perspectives is a crucial starting point in making AI systems more responsible or trustworthy.

Now, let's start to think about some of these ethical issues with AI. We will focus on generative AI, which is the type of AI that needs a foundation model to work, and that requires lots of data to create. We can categorise ethical issues in AI according to the five layers of AI which we introduced earlier in the course. These were:

- the engineering of AI
- methods of AI
- applications of AI
- governance of AI, and
- narratives of AI.

Engineering of AI

Firstly, within ‘engineering of AI’, we have the issue of **human labour**. We often discuss AI taking people's jobs. But we should also focus on the poor conditions for workers who help train AI systems, like ChatGPT, through online microtasking. This method is common in AI development, and can lead to unfair and harmful work conditions.

Secondly, is the issue of **climate change**. Creating foundation models consumes a huge amount of energy and requires cooling systems for the servers. This not only affects the environment, but sometimes involves the use of materials obtained through unethical means.

Methods of AI

The next category is ‘methods of AI’. The first issue here is **bias**. You may have seen examples of biased and harmful text or images that foundation models can produce. For example, if you ask for an image of a computer science professor, you will probably not get an image of someone that looks like me.

The second issue is **the internet as a data source**. Although some of these biases can be fixed, it doesn’t change the underlying inequality in using data from the internet. Even though there are about seven thousand languages worldwide, most internet content is in just ten languages. US English dominates AI, and this affects education. It’s important to have AI tools that respect different cultures to make learning more relatable.

The next issue is **privacy**. Has appropriate copyright and informed consent always been secured to use data in training models? And are there robust procedures to prevent unauthorised access and data leaks? And where does data entered into corporate-owned services go?

Next, **open source**. Many scientists argue that letting the wider community of computer scientists freely test and improve AI can make it more reliable. But companies creating foundation models aren’t allowing this.

Finally, within the ‘methods of AI’ category is **decolonial AI**. These issues around privacy, opaqueness of foundation models, underrepresentation of diverse people in training data sets, and people’s lack of control over their own data, highlight how foundation models may concentrate power and homogenise culture. Many argue that these AI methods reflect colonial legacies.

Applications of AI

Now let’s look at ‘applications of AI’. The main issue here is **misuse**. Misusing generative AI can lead to the spread of misinformation, cyberbullying and cyber attacks, and the creation of deepfakes that can undermine trust and manipulate public opinion. A more specific example of misuse in the educational context is **academic dishonesty**. Generative AI can be used to plagiarise essays, fabricate research data or code, and automate coursework solutions.

Governance of AI

The next category is ‘governance of AI’. A big difficulty with governing AI is that it’s hard to know what **unintended consequences** can emerge when AI moves out of research labs and into our health, education or transport systems.

Also, **who gets to make decisions** about AI, especially when it affects us. Should it be the tech companies, our governments, or should we also get a say in how AI is governed?

Here we've highlighted a variety of ethical issues that have come up in relation to AI, and some common themes in AI ethics that have emerged.

Next, I recommend that you try to seek out different perspectives on these issues, in your conversations with friends or colleagues. Use these conversations to find your own voice, and to better understand where you sit on a spectrum of views. This should be your starting point.