Teaching ethics to AI developers

Maurice Chiodo
mcc56@cam.ac.uk
Centre for the Study of Existential Risk,
University of Cambridge

The Ethics in Mathematics Project:
https://www.ethics.maths.cam.ac.uk

27 June 2023

1. Introduction

This talk is about what aspects of mathematics we need to teach to mathematicians who might go and work in AI. However, it covers aspects that indeed should be taught to all mathematicians.

I will have *ideas slides* (with not much text), and *content slides* (with a lot of text). If you'd like a copy of the slides, or have any other questions, please email me.

Some initial thoughts:

- How do we address bias in algorithms, when our students don't believe that mathematics can be "biased"?
- How do we address harmful outcomes of mathematical work, such as AI, when our students don't believe that mathematics can cause harm?
- How do we address totally false mathematical outputs, when our students apply mathematics in an unfalsifiable way?
- How can we teach the politics and societal impact of AI, when our students believe that all mathematical technologies are neutral?

Our mathematics students are part of, and often lead, the teams, start-ups and companies building AI. They are changing the world with little ethical training, or understanding of what their work does to people.

We seldom address this as mathematics educators. Instead, we often simply teach mathematics for those few percent who end up in academic mathematics.

2. Common responses to Ethics in Mathematics

I've taught Ethics in Mathematics (EiM) for over 7 years now. Here is a list of the most common replies that I have received by students, colleagues and superiors when in such discussions:

- Mathematics is pure.

- Mathematics is pure.
- Ethics is just an opinion; it is not mathematically precise.

- Mathematics is pure.
- Ethics is just an opinion; it is not mathematically precise.
- This is not my problem.

- Mathematics is pure.
- Ethics is just an opinion; it is not mathematically precise.
- This is not my problem.
- This is not my job.

- Mathematics is pure.
- Ethics is just an opinion; it is not mathematically precise.
- This is not my problem.
- This is not my job.
- Who's ethics?

- Mathematics is pure.
- Ethics is just an opinion; it is not mathematically precise.
- This is not my problem.
- This is not my job.
- Who's ethics?
- This is anti-progress/finance/capitalism.

- Mathematics is pure.
- Ethics is just an opinion; it is not mathematically precise.
- This is not my problem.
- This is not my job.
- Who's ethics?
- This is anti-progress/finance/capitalism.
- I'm just doing my job, what I was told to do.

- Mathematics is pure.
- Ethics is just an opinion; it is not mathematically precise.
- This is not my problem.
- This is not my job.
- Who's ethics?
- This is anti-progress/finance/capitalism.
- I'm just doing my job, what I was told to do.
- If I don't do it, someone else will.

- Mathematics is pure.
- Ethics is just an opinion; it is not mathematically precise.
- This is not my problem.
- This is not my job.
- Who's ethics?
- This is anti-progress/finance/capitalism.
- I'm just doing my job, what I was told to do.
- If I don't do it, someone else will.
- What qualifies a mathematician to speak about ethics in mathematics if they are not a philosopher?

- Mathematics is pure.
- Ethics is just an opinion; it is not mathematically precise.
- This is not my problem.
- This is not my job.
- Who's ethics?
- This is anti-progress/finance/capitalism.
- I'm just doing my job, what I was told to do.
- If I don't do it, someone else will.
- What qualifies a mathematician to speak about ethics in mathematics if they are not a philosopher?
- What qualifies a philosopher to speak about ethics in mathematics if they are not a mathematician?

- I see no other mathematicians doing this, so why should I?

- I see no other mathematicians doing this, so why should I?
- Ethics wastes time that could be used for more mathematics.

- I see no other mathematicians doing this, so why should I?
- Ethics wastes time that could be used for more mathematics.
- I just do mathematics, and do so because I don't want to think about anything else.

- I see no other mathematicians doing this, so why should I?
- Ethics wastes time that could be used for more mathematics.
- I just do mathematics, and do so because I don't want to think about anything else.
- It is my right to do whatever mathematical work I want.

- I see no other mathematicians doing this, so why should I?
- Ethics wastes time that could be used for more mathematics.
- I just do mathematics, and do so because I don't want to think about anything else.
- It is my right to do whatever mathematical work I want.
- Including social and ethical considerations means we're no longer solving a mathematics problem.

- I see no other mathematicians doing this, so why should I?
- Ethics wastes time that could be used for more mathematics.
- I just do mathematics, and do so because I don't want to think about anything else.
- It is my right to do whatever mathematical work I want.
- Including social and ethical considerations means we're no longer solving a mathematics problem.
- This is not PhD-level mathematics, so it's not actually mathematics.

- I see no other mathematicians doing this, so why should I?
- Ethics wastes time that could be used for more mathematics.
- I just do mathematics, and do so because I don't want to think about anything else.
- It is my right to do whatever mathematical work I want.
- Including social and ethical considerations means we're no longer solving a mathematics problem.
- This is not PhD-level mathematics, so it's not actually mathematics.
- It's never the mathematics; it's how you use it.

- I see no other mathematicians doing this, so why should I?
- Ethics wastes time that could be used for more mathematics.
- I just do mathematics, and do so because I don't want to think about anything else.
- It is my right to do whatever mathematical work I want.
- Including social and ethical considerations means we're no longer solving a mathematics problem.
- This is not PhD-level mathematics, so it's not actually mathematics.
- It's never the mathematics; it's how you use it.
- ok, I agree this is important; can you please give me the axioms and deterministic algorithm for ethics.

Hint:

Don't get up in front of a room full of mathematicians and talk about Ethics in Mathematics unless you have a response prepared for each of the statements or questions above.
(I can provide this, if you want).

3. Why does this happen?

We have seen this before in history, where a group of people who were highly skilled and highly trained in an advanced technology did not have a full awareness of how to "behave" when using their skills.

These were the longbow archers of Henry V.

In the context of mathematics teaching:

- We train our students to be subservient by saying "Find x", and thus they learn to never ask "Why?". And we often don't teach them how to *choose* problems based on non-mathematical metrics (e.g., societal relevance, urgency, etc.).
- We train our students to strip away the "fuzzy" or "messy" stuff to get down to the "real mathematics", and thus they learn to ignore non-numerical aspects. Many of the exercises we give them are artificially simple and clean in order for them to focus on the one "mathematical lesson" that it is supposed to teach them. Reality is not clean.
- We teach our students the *maximum amount of mathematics possible*, and thus by exclusion they learn very very little else. This is a particular problem in countries that don't have general education requirements, or minor subjects as part of their undergraduate education (e.g., the UK).

- We train our students in logical and mathematical precision, and thus they learn to disregard imprecise things and never learn to navigate murky waters challenging real problems.
- They do all this for thousands of hours - a student completing a 4 year mathematics degree might do nearly 10,000 hours of mathematics - often at the expense of extra-curricular interests. So even saying "learn it on your own" doesn't work!
- Our current teaching changes the way they think and behave, and the way they see the world and its problems. For better, and for worse.

4. The three pillars of teaching Ethics in Mathematics

To teach Ethics in Mathematics *effectively*, you need to do 3 key things:

1. Deliver a dedicated, labelled course on Ethics in Mathematics.
2. Integrate Ethics in Mathematics across the curriculum.
3. Create an environment whereby your colleagues support discourse around Ethics in Mathematics, or at the very least do not damage the endeavour of teaching it by sabotaging it in front of students.

These are ordered in decreasing financial cost, but increasing political cost.

Some resources.

Article on how to teach Ethics in Mathematics:
Teaching Ethics in Mathematics, M. Chiodo, P. Bursill-Hall, LMS Newsletter **485**, 22-26, November 2019.

Recorded lecture course on Ethics in Mathematics:
https://ethics.maths.cam.ac.uk/course/lectures
(A set of accompanying lecture notes is forthcoming)

Mathematical exercises with an ethical component:
https://ethics.maths.cam.ac.uk/course/
(A much larger set is forthcoming)

5. Upskilling the population and the workers;

an aside on what we migh need to teach the rest of the population

Society started teaching mathematics more broadly as it became more necessary during each stage of industrialisation. People *needed* more mathematics to do work.
The type of mathematics chosen reflected the work a lot of people needed to do. This has happened in each of the first three industrial revolutions:
1st industrialisation: Hand production to machines (1760-1820s).
2nd industrialisation: The technological revolution (1871-1914).
3rd industrialisation: The digital revolution (1950-2000).
4th industrialisation: Automation of the economy through cyber-physical systems such as AI (*forthcoming*).

Each time we enter a more technical world, people need **MORE** mathematical skills, not fewer (though possibly different ones).

It isn't clear exactly what those are yet for the AI revolution. But it is not the 12x tables...

This will be somewhat like learning to use computers in the digital revolution. But it will definitely require *upskilling* as the AI systems get more complex, not downskilling. This is perhaps counterintuitive, given that "Now ChatGPT writes the article for me".

6. Ethics in Mathematics helps you do better mathematics!

We produced a tool to help mathematicians carry out their work responsibly and ethically. It examines the typical workflow of the mathematical development process and identifies some of the of key questions that need to be asked or addressed at each stage:

1. Deciding whether to begin.
2. Diversity and perspectives.
3. Handling data and information.
4. Data manipulation and inference.
5. The mathematisation of the problem.
6. Communicating and documenting your work.
7. Falsifiability and feedback loops.
8. Explainable and safe AI.
9. Technological artefacts have politics.
10. Emergency response strategies.

This was a joint work with Dennis Müller:
Manifesto for the Responsible Development of Mathematical Works A Tool for Practitioners and for Management. arXiv:2306.09131

We're hiring!

We're hiring!

(but we have no money)

We're hiring!

(but we have no money)

If you'd like to learn about, make use of, or collaborate on, any part of our project, please do get in touch!
Current collaborations include:

- Developing additional (open-access) teaching resources for teaching Ethics in Mathematics.
- Writing domain-specific supplementary documents for our general manifesto on responsible development.
- Writing papers with legal scholars on novel regulatory techniques for AI.
- Leading an interview-based qualitative study on the values and mindsets of (academic) mathematicians.