AI and Society Seminar 10
How are large language models (LLMs) impacting education in NZ, and how should NZ teachers and policymakers respond?

Today's discussion

- Kit Willett (Selwyn College, Auckland)
- Kathleen Kaveney (Post-Primary Teachers' Association)
- Robyn Caygill (Chief Education Advisor, Ministry of Education)
- Neil Miller (NZQA)
- Simon McCallum (Engineering and Computer Science, VUW)

Some general themes for discussion

Some general themes for discussion

There should be good communication between teachers, students, AI researchers, educational policymakers and AI regulators.

Some general themes for discussion

There should be good communication between teachers, students, AI researchers, educational policymakers and AI regulators.

- Information should flow from teachers and students, about what works and what doesn't, and what additional things would be helpful.

Some general themes for discussion

There should be good communication between teachers, students, AI researchers, educational policymakers and AI regulators.

- Information should flow from teachers and students, about what works and what doesn't, and what additional things would be helpful.
- Information should flow from AI researchers about what technology is available, and what is on the way.

Some general themes for discussion

There should be good communication between teachers, students, AI researchers, educational policymakers and AI regulators.

- Information should flow from teachers and students, about what works and what doesn't, and what additional things would be helpful.
- Information should flow from AI researchers about what technology is available, and what is on the way.
- Information should flow from educational policymakers to teachers, and to AI regulators.

Some general themes for discussion

There should be good communication between teachers, students, AI researchers, educational policymakers and AI regulators.

- Information should flow from teachers and students, about what works and what doesn't, and what additional things would be helpful.
- Information should flow from AI researchers about what technology is available, and what is on the way.
- Information should flow from educational policymakers to teachers, and to AI regulators.

NZ is small and tech savvy enough that we should be able to get the structures right!

Ali's hobby horse: detection mechanisms

Ali's hobby horse: detection mechanisms

I co-lead a project on social media governance at the Global Partnership on AI.

- We've thought a lot about the risks of LLMs in social media, and how to address these.

Ali's hobby horse: detection mechanisms

I co-lead a project on social media governance at the Global Partnership on AI.

- We've thought a lot about the risks of LLMs in social media, and how to address these.
- LLMs can readily produce disinformation. An automated detector for LLM-generated content would be very useful in identifying this.

Ali's hobby horse: detection mechanisms

I co-lead a project on social media governance at the Global Partnership on AI.

- We've thought a lot about the risks of LLMs in social media, and how to address these.
- LLMs can readily produce disinformation. An automated detector for LLM-generated content would be very useful in identifying this.
- Of course, detection may be useful in many other domains too, including education.

Ali's hobby horse: detection mechanisms

I co-lead a project on social media governance at the Global Partnership on AI.

- We've thought a lot about the risks of LLMs in social media, and how to address these.
- LLMs can readily produce disinformation. An automated detector for LLM-generated content would be very useful in identifying this.
- Of course, detection may be useful in many other domains too, including education.
- Our group is proposing a new piece of regulation about detection mechanisms.

Our proposed rule

We propose:

- Any company developing a state-of-the-art FM must demonstrate a reliable detection mechanism *as a condition of release*.

Our proposed rule

We propose:

- Any company developing a state-of-the-art FM must demonstrate a reliable detection mechanism *as a condition of release*.

Why must the detector be in place *prior to release*?

Our proposed rule

We propose:

- Any company developing a state-of-the-art FM must demonstrate a reliable detection mechanism *as a condition of release*.

Why must the detector be in place *prior to release*?

- FMs must be *instrumented* to support detection. Detection won't be possible just by inspecting the generated content.

Our proposed rule

We propose:

- Any company developing a state-of-the-art FM must demonstrate a reliable detection mechanism *as a condition of release*.

Why must the detector be in place *prior to release*?

- FMs must be *instrumented* to support detection. Detection won't be possible just by inspecting the generated content.

We're discussing this proposal with EU regulators, and at OECD.

Today's discussion

- Kit Willett (Selwyn College, Auckland)
- Kathleen Kaveney (Post-Primary Teachers' Association)
- Robyn Caygill (Chief Education Advisor, Ministry of Education)
- Neil Miller (NZQA)
- Simon McCallum (Engineering and Computer Science, VUW)