

Are LLMs Weapons of Math Destruction?

Recall the WMD Framework:

According to O'Neil, a Weapon of Math Destruction has three key features:

1. **Opacity** – The system is difficult to understand or challenge
2. **Scale** – It affects large numbers of people
3. **Damage** – It causes harm, especially to vulnerable groups

Often includes **feedback loops** that reinforce inequality

Work with a partner. Consider large language models (ChatGPT, Gemini, Claude, etc.). For each criterion, decide whether LLMs currently meet it and explain your reasoning.

1. Opacity

Are LLM's opaque?

2. Scale (Where are LLMs being used today at a large scale?)

3. Damage: Discuss possible harms:

4. Feedback Loops: Can LLM's create self-reinforcing cycles?

If you're not sure what is meant by that, consider AI-generated content being used to train future models, or perhaps students over-relying on AI leading to less skill development.

Scenario Analysis:

Choose one of the two scenarios below, and determine if it constitutes a “WMD”.

Scenario A: AI Writing Evaluation

A university uses an LLM to score student essays.

Questions:

- What might the model reward or penalize?
- Who might be disadvantaged?
- Would students be able to challenge the score?

Scenario B: AI Customer Service

A company replaces human support with an LLM chatbot.

Questions:

- Who benefits?
- Who might struggle?
- What happens if the system gives incorrect information?