Mahveen Raza
High School Student
HWDSB, Hamilton, Canada
mraza3580@hwdsb.on.ca
August 13, 2025

To: Education Materials Chairs, NeurIPS 2025
Subject: Submission to the NeurIPS 2025 Call for Education Materials

Dear Education Materials Chairs,

I am pleased to submit our educational resource, *How AI Vaccines Work: Training, Defenses, and Testing to Keep AI Safe*, to the NeurIPS 2025 Education Track. The resource is designed for non-expert audiences and aligns with the call's goals of accessible, engaging, and reusable AI education materials.

Audience & Engagement
Target audience: high school students, undergraduate students in non-CS majors, educators, and the general public with no prior AI background.
Expected engagement time: 10 minutes (visual explainer + interactive visualization + blog-style explainer; optional 5-minute video).

Overview
The material explains the concept of an "AI Vaccine" that provides layers of defence against prompt injections that probe harmful content or misinformation from an AI model. It uses the analogy of human vaccines and AI vaccines to explain the concept in clear, non-technical terms. Using curated examples of well-known AI models' responses, we combined them into an interactive visualizer that allows exploration of different prompt defence techniques and their effectiveness against different types of malicious prompts. The goal is to let a non-technical audience understand AI's vulnerability to malicious prompts, promoting critical thinking about cybersecurity and safe online use of AI, while remaining approachable and classroom-ready.

Submission Components

- Infographic: `infographic.pdf`

- Short Video Explainer: overview of concepts and key takeaways (`video explainer.mp4`, 5 minutes)

- Blog-Style Document: overview of concepts and key takeaways (`blog explainer.pdf`, ∼3 minutes read time)

- Interactive Visualization: ai-vaccines.vercel.app

- GitHub Repository (source code & dataset): github.com/maximus-powers/ai-vaccine-experiment

Educational Value & Reusability
The resource is concise, modular, and easily adoptable in lessons, clubs, and outreach. It includes transparent references, open access to data and code, and clear descriptions educators can adapt to different time budgets and learner levels.

Authorship
This material has been prepared with and for high school learners to promote AI literacy and climate awareness.

Thank you for your consideration. I would be delighted to share this resource with the NeurIPS community and support its dissemination to educators and students.

Sincerely,
Mahveen Raza