Introduction

Project Frozone involves:

- Stephen Davies (professor of Computer Science & Data Science)
- Bethanie Hackett (Computer Science & Political Science senior)
- Garrett McKenzie (Computer Science & Data Science junior)
- Laur Rider (Computer Science & English junior)

Our goal is to determine whether **AI chatbots** can be deployed to help **mitigate political polarization** in a society.

We're training a **Large Language Model** (LLM) called "**Frobot**" to detect instances of **misinformation**, **toxic speech**, and **entrenched bias** when they appear in online political conversation, and to **counteract** those elements with its responses.

Objectives/Outcomes

We plan to evaluate Frobot in two ways:

1. *Deploying it to a chatroom with human participants.* Experimental subjects will engage with Frobot (and other vanilla bots) without knowing they are bots, and produce chat log transcripts for us to analyze using the scientific techniques of **Discourse Analysis** and **Qualitative Coding**. Participants will also complete **surveys** in which they rate the degree to which the bots were effective communication partners, promoted healthy dialogue, and the like.

2. *Simulating Frobot deployment on a larger scale.* Using the scientific technique of **Agent-Based Modeling** (ABM), we will design a computer simulation of a large, **virtual social network**, in which both simulated humans and Frobots interact. By varying features of the network – including the ratio of humans to Frobots, the patterns by which individuals form "friendships" with other individuals, and the frequency with which individuals interact – we will gain understanding about what **aggregate effect** Frobots might have on society at large.

Budget

We seek funding for two things:

1. *Cloud compute time and storage space to train Frobot.* Training LLMs on large amounts of text data is expensive. We plan to use the Google Gemini and Vertex AI suite of tools, at an estimated cost of **$1500**.

2. *Computational power to simulate the ABM.* Analyzing the behavior of an ABM requires running it a large number of times, both to "even out" the randomness of each individual trial, and also in order to vary the simulation's inputs to discover how various parameter settings influence the aggregate results. We will use Google Virtual Machines to carry out these resource-heavy computations, and estimate **$1000** for this.

Summary:

1. Training resources: $1500
2. Simulation resources: $1000
- **Total funding request: $2500**