

# KARINA BELOKPOV

Cambridge, MA | karina-belokapov@outlook.com

## EXPERIENCE

### SERI MATS

Summer 2023

#### Stanford Existential Risk Institute ML Alignment Theory Scholar

- Created a demo of consumer AI capabilities within social media, mentored by Jeffrey Ladish (former CISO for Anthropic)
- Designed and implemented experiment to gauge interaction ratios and trust-building abilities of an AI agent in a real-world scenario (Discord)
- Research presented to the Biden Administration

### ITSTEP Academy

2022-2023

#### STEM instructor

- Designed and taught curriculum to 8-11 year olds, covering programming, operating systems, networking, robotics, control theory, security, machine learning, etc

## EDUCATION

### Bellevue College

2023-2024

#### Running Start - Multivariable Calculus, Differential Equations, Real Analysis

- Presented at math and physics colloquium run by professors for professors about the state of computational neuro
- Multivariable Calculus, Differential Equations

### Interlake High School

2021-2023

#### International Baccalaureate Advanced Learning Program

- Ran Go Club (organized tournaments, designed variants of the game)
- FIRST robotics: Programming mentor, Java
- IB Physics HL2, IB Biology SL2, IB Apps&Interps HL2 (Set Theory, Real Analysis, Logic)

## PROJECTS

### EEG Research (Emergent Ventures grant)

2023-2024

Received an \$8,000 grant from Tyler Cowen to **replicate EEG experiments** and evaluate their utility as brain-computer interfaces in the real world. Created a new low-friction **data collection framework** for the OpenBCI headset, analyzed data in Python using classical signal processing and Variational Autoencoders, attempted to replicate neural entrainment paper.

### Automation, Scraping, Data Analysis

2019-Present

- Built an app that correlates episodes of idiopathic dizziness with local pressure changes (using NOAA barometric data)
- Built a tool to assist students in choosing their degree at the University of Washington based on the intersection of their interests and course requirements. Wrote a scraping pipeline in Python to grab classes, parse and visualize prerequisite graphs based on interests, & automatically color-code majors.
- Analyzed trends in security** based on keyword frequency over time in MITRE CVE database and Bleeping Computer articles
- Created automatic data scraping agent using GPT-4 that takes in elements of any webpage and scrapes all following the given pattern.

### Machine Learning Projects

2021-Present

- Fine tuned** llama-70B on chat logs on an A100 to demonstrate risks associated with high quality personality/style emulation
- Performed Direct Preference Optimization, a tuning technique, on a GPT-Neo-125M model and mapped out how that changed the weight topology with matplotlib
- Built a full stack webapp to non-deterministically ZKP (Zero Knowledge Proof) **pieces of gossip** between N parties
- Scraped and preprocessed a dataset of thousands of images and consistent, curated ratings from r/truerateme, **trained an SVM** to rate people's facial features & created a streamlit website to host it

### Hardware

2021-Present

- Designed and built **galvanic skin sensor** for measuring emotional arousal from skin resistance
- (Created **RFID sign in system** for Robotics team. Bought and assembled hardware, wrote Python and Arduino code, tested and maintained. Also did generic **CAD** (Onshape) and **shop work** (CNC, bandsaw, etc) for various parts of the robot.

## HONORS

- Atlas Fellow
  - \$10,000 Scholarship and Rationality Camp (~2% acceptance rate)
- Recurse Center Alumni
  - Programming Retreat
- HAM Radio Technician License
  - I have access to UHF and VHF bands
- Taco Live Mas Scholar
  - \$10,000 award for discussing optogenetics
- 99th+ percentile contributor on the security stackexchange for around a year

## SKILLS

### Programming

- Python, MATLAB, C, C++, R, Rust, Golang, HTML/CSS, React.js, Node.js, three.js, SQL, Jupyter, vim
- Misc. softwares (I learn very fast)

### Mathematics

- Real Analysis, Discrete Math, Mathematica

### Data analysis

- pandas, numpy, matplotlib

### Unix Systems Administration

- Networking
- Security

### Design

- Fusion360, Onshape, Blender, Figma, Alphafold