

Alexander Karpekov

alex.karpekov@gmail.com | +1 (650) 660-7729 | U.S. (Permanent Resident) | LinkedIn: [/karpekov](#) | alexkarpekov.com

Political Science and Economics student turned Big Tech guy turned Computer Science grad student. After spending almost ten years as a Data Scientist in the industry (seven years at Google, and three years at a startup), I decided to go back to school to pursue a PhD in Computer Science with a focus on AI and Machine Learning.

I have a particular interest in the field of Explainable Artificial Intelligence, and the intersection of AI and social sciences. I am also always looking to explore how to use data visualizations for storytelling.

EDUCATION

Georgia Institute of Technology

2024 - current | Atlanta, GA

PhD in Computer Science

Co-advised by Prof. [Sonia Chernova](#) and Prof. [Thomas Plötz](#)

First year PhD students with a concentration on Human Behavior Recognition from Sensor data and Explainable AI.

Georgia Institute of Technology

2024 | Atlanta, GA | 3.9/4.0

MS in Computer Science

Machine Learning Specialization

Completed 2nd Master's Degree remotely while working full time at Google.

University of California, San Diego

2015 | San Diego, CA | 3.8/4.0

MA in Economics

Worked as a Teaching Assistant for 3 graduate-level classes in Statistics and Econometrics for 120+ grad students.

Moscow State University

2013 | Moscow, Russia | 92/100

BA in Political Science

SKILLS

Programming (in order of proficiency):

Python, SQL, TypeScript, R, Stata, C, Java.

ML & Data Science: PyTorch, Hugging

Face, TensorFlow + Keras, LangChain, LanceDB, Scikit-learn, Statsmodels, XGBoost.

Data Analysis and Visualization:

NumPy, Pandas, SciPy, Jupyter, Colab, Matplotlib, Altair, Plotnine.

Front End: Svelte, D3, Tailwind CSS,

Figma, Adobe Illustrator.

HOBBIES

Rowing (GeorgiaTech Crew), CrossFit (8 years), Snowboarding, Rock Climbing.

EXPERIENCE

GEORGIA TECH | Researcher

2023 - Present | Atlanta, GA

Automated Human Activity Recognition: Developing a methodology to automatically detect, analyze, and cluster human activity using in-house sensor data. Working with Professor Sonia Chernova.

- Pre-trained BERT model on sensor sequences data using mask-language modeling, and fine-tuned for the clustering task using SCAN loss. Used this model to create and cluster embeddings to identify groups of spatial and temporal human activities.
- Created cluster and sensor sequence interpretability tool using D3.

Transformer Explainer: Developed an interactive educational tool to explain the inner workings of the Transformer models. Working with Professor Polo Chau. The tool is available [here](#). Paper is here: [arxiv](#).

- Built a web-based, interactive tool that demonstrates the functionality and architecture of GPT-2, including in-browser next-token prediction and temperature-adjusted sampling. Used Svelte Kit and D3.
- We won the [best poster award](#) at IEEE Visualization conference in 2024.

GOOGLE | Senior Data Scientist (L5) @ YouTube and Google Search

2017 - 2024 | San Francisco, CA and Dublin, Ireland

Summary: Worked as a Data Scientist in Google Search and YouTube Music, with the main focus on A/B experiment design and evaluation to improve search results quality and music recommendation algorithm.

Core Expertise: Statistical analysis using A/B testing and causal impact methodologies like propensity score matching; Clustering and classifications tasks; Embedding space construction; User data analytics and visualization, managing datasets with billions of entries.

Project Highlights:

- Developed a pathfinding algorithm in song embedding space, improving music recommendations that led to 3% boost in user engagement and music discovery rates.
- Implemented a new methodology to cluster YouTube multi-billion music corpus using text, sound, search, and co-watch embeddings, which led to a 30% reduction in harmful watchtime and a 0.5% increase in music revenue (\$100s millions).
- Created a new counterfactual causal impact methodology to evaluate the impact of the new feature launch on user engagement and conversion that helped establish no statistically significant long-term effects on key business metrics. The analysis was instrumental to halt the global rollout at Engineering and Product VP-level.

DATAMINR | Data Analyst

2015 - 2017 | New York, NY and London, U.K.

- Built statistical models to automatically classify Twitter user handles.
- Conducted Twitter user clustering and unsupervised learning using networks analysis methodologies to improve news discovery algorithms.