# Alexander **Karpekov**

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

**Future**: Work in AI and ML research & development space with a focus on model interpretability and explainability, and causal impact. Join a dynamic, synergistic team that values diverse strengths and has a shared vision for impact and success. Looking for in-office or hybrid opportunities, I believe in the importance of interpersonal connections. Preferred location: NYC.

**Past**: Data Scientist and ML researcher with 10 years of Big Tech and startup experience. Completed second Master's in CS while working full time at Google and being promoted twice. Studied and worked in 9 cities across 5 countries: San Francisco, New York, San Diego, Atlanta (U.S), Dublin (Ireland), London (U.K.), Strasbourg (France), Moscow, and St. Petersburg (Russia).

## EDUCATION

### Georgia Institute of Technology
2024 | Atlanta, Georgia | 3.9/4.0

**MS in Computer Science**
**Machine Learning Specialization**

Completed 2nd Master's Degree remotely while working full time at Google. Took a semester-long sabbatical to take classes and conduct research on Atlanta campus.

### University of California, San Diego
2015 | San Diego, California | 3.8/4.0

**MA in International Economics**

Worked as a Teaching Assistant for 3 graduate-level classes in Statistics and Econometrics, leading sessions for 120+ students. Received the best TA award.

### Moscow State University
2013 |Moscow, Russia | 92/100

**BA in Political Science**

## SELECTED COURSEWORK

Deep Learning for Text
Generative Artificial Intelligence
Explainable Artificial Intelligence
Reinforcement Learning
Computer Vision
Advanced Econometrics
Survey Design and Evaluation
GIS and Spatial Data Analysis

## SKILLS

**Deep Learning**: pytorch, tensorflow, keras, huggingface, sklearn, statsmodels, xgboost
**Python**: numpy, pandas, scipy, jupyter, colab, matplotlib, altair, plotnine
**Programming** (in order of proficiency): Python, SQL, R, Stata, C++, Java, Javascript
**Languages:** EN, RU, FR, DE, ZH

## HOBBIES

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

## EXPERIENCE

### **GOOGLE** | Senior Data Scientist @ YouTube and Google Search
2017 – Present | San Francisco, California and Dublin, Ireland

**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.
- Created a novel Transformer-based methodology to embed user action sequences on YouTube music app, which informed YouTube-wide product strategy for 2024.
- 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.
- Led company-wide effort for reporting automation using Python instead of Excel.

## PROJECTS

**Transformer Block Visualizer**
2024 @ GT

*[Ongoing]* Developing a Transformer block interpretability tool for educators and ML developers that sheds light on the Transformer training process.

**Gradient-Attention for Feature Importance Extraction in Transformers** | 2023 @ GT

Created a feature extraction tool for Transformer-like models (e.g., BERT and BART) for sentiment analysis of text data.

**LLM Sensitivity to Prompt Perturbations**
2023 @ GT

Created a statistical tool to evaluate LLM (GPT3, GPT4 and Llama 2) sensitivity to various prompt perturbations, such as lexical, semantic, structural, and tonal changes in the LLM input.

**Reinforcement Learning Model Interpretability** | 2023 @ GT

Created a Q-value evaluation technique for Lunar lander RL environment to assess the agent learning progress throughout training.

**Using Stock Market Data to value Political Connections** | 2015 @ UCSD

Created a methodology to estimate the impact of international sanctions on the Russian companies' market value using stock data and Capital Asset Pricing Model.

**Policy Impact Evals using Satellite Data**
2014 @ UCSD

Used nightlight brightness data and GIS tools to approximate GDP growth in China. Ran geographically-weighted regression to evaluate a 15% increase in GDP as a result of the province-level policy in 2008-2013.