

Alexander Quispe

PH.D STUDENT IN COMPUTER SCIENCE • CALTECH

438 S Catalina Ave, Pasadena, CA 91106

✉ +1 (202) 912-0242 | ✉ aquisper@caltech.edu | 🗂 alexanderquispe.github.io | 📡 alexanderquispe

Education

California Institute of Technology

PH.D STUDENT IN COMPUTING AND MATHEMATICAL SCIENCES

Pasadena, CA

September 2025 - 2029

Cornell University

M.S. COMPUTER SCIENCE

New York City, NY

September 2024 - May 2025

Massachusetts Institute of Technology

VISITING STUDENT

Boston, MA

2019 - 2020

Ludwig Maximilian University of Munich

MSc. QUANTITATIVE ECONOMICS - PHD TRACK

Munich, Germany

2018 - 2020

Pontificia Universidad Católica Del Perú

BS ECONOMICS

Lima, Peru

2013 - 2017

Visiting Position

2019-2020 **Visiting Scholar at Harvard University**, Laboratory for Innovation Science. **Host Prof. Karim R. Lakhani**

Experience

GitHub Research Team

RESEARCHER, INDUSTRY CLASSIFICATION PROJECT

Remote / USA

2024-2025

- Designed and implemented a **semantic search pipeline** using embeddings to match 511k GitHub repositories to NAICS industry codes.
- Developed a **GPT-4 agent validation framework** to confirm industry alignment, scoring repositories and filtering ~2,500 high-confidence matches.
- Fine-tuned state-of-the-art **Transformer models** (RoBERTa, ModernBERT, DeBERTa) on GitHub-NAICS data, achieving up to **86.45% F1-score**.
- Applied the system to **global inference analysis** (USA, Germany, Brazil, Rwanda, etc.), mapping open-source contributions across income levels.

Cornell Tech, Cornell University

GRADUATE RESEARCHER

New York, NY

2024-2025

- Replicated core results from the paper **“More Robust Doubly Robust Off-policy Evaluation”** (Jiang et al., 2020), as part of coursework and research on reinforcement learning and offline evaluation methods.
- Developed a full simulation pipeline in Python to compare the performance of various off-policy evaluation (OPE) estimators: **Importance Sampling (IS)**, **Doubly Robust (DR)**, **Marginalized IS (MIS)**, **Direct Method (DM)**, and **Double Reinforcement Learning (DRL)**.
- Utilized **GPU acceleration** to speed up the Monte Carlo simulation process across different sample sizes. Conducted over 200 replications per setting.

The World Bank

Washington, DC

2020-2025

RESEARCHER

- Led research on the **impact of generative AI (ChatGPT) on software development activity** using GitHub administrative data, applying econometric and machine learning methods.
- Taught a graduate-level **Causal Machine Learning course** (causal trees, causal forests with Uber’s CausalML).
- Developed **AI-driven tools** including EconMentor (GPT-4 for research question generation) and DigitalPillarAI (semantic analysis and classification of project documents).
- Created an open-source **OSRM Python package** for real-time driving distance estimation, integrating geospatial ML workflows.
- Applied **satellite imagery, Earth Engine, and Google Open Buildings** to build raster-based ML indicators for crisis and poverty analysis.
- Contributed to the **Data Skills Certification Program**, designing modules on AI/ML applications for development research.

MIT Sloan School of Management

Cambridge, MA

2019-2020

RESEARCH ASSOCIATE

- Applied **machine learning and advanced econometric methods** (Quantile, Poisson, IV Poisson with bias correction) to large-scale trade and communication datasets.
- Developed and analyzed **geospatial and historical datasets** (cotton trade flows, containerization, international shipping records) to study how technology adoption reduces communication and transportation costs.

Max Planck Institute for Innovation and Competition

Munich, Germany

2018-2019

RESEARCH ASSISTANT

- Developed an algorithm using **Google Directions API to calculate commuting times** between residence and workplace municipalities for inventors by train, car, or bicycle in Python.
- Conducted nonparametric regressions to estimate the probability of inventors staying in a district after tax increases and used **Amazon Web Services for parallel computing** to calculate confidence intervals with Bootstrap in R.

Department of Economics, LMU Munich

Munich, Germany

2018-2019

RESEARCH ASSISTANT

- Applied **causal inference (RDD)** with geocoded school data to study education policy impacts on labor outcomes.
- Used **lasso/ridge regression** to analyze gender representation, finding stronger causal effects of female mayors on women’s political advancement.

- Estimated the effectiveness of cash bonuses for retaining public teachers in remote areas of Peru (2015–2018) using regression discontinuity design (RDD).
- Built a comprehensive teacher migration database and developed a **fuzzy matching algorithm (Levenshtein distance)** to merge local data with a national roster of four million teachers.

Statistical software and AI Projects

csdid.py - Difference-in-Differences including Double Robust Estimation-with **Pedro H. C. Sant'Anna** - **212,000 downloads (PyPI)**

synthdid.py & Synthdid.jl - Python and Julia implementation of Synthetic difference in differences method based on Athey et al. (2021) - **15,200 downloads (PyPI)**

TreatmentEffectRisk.py - Treatment Effect Risk: Bounds and Inference-with **Nathan Kallus**

HDMJL.jl - Julia Package of high dimensional econometrics - with **Victor Chernozhukov**

Sensemakr.jl - Julia package for sensitivity analysis tools based on Cinelli et al. (2020)

osrm.py - Python package to calculate driving distances for free using OSRM engine

ILLA - AI-trained virtual assistant skilled in identifying potential obstetric and gynecological violence-**Media**.

DigitalPillarAI - AI tool integrated with GPT4 engine to classify Project Appraisal Documents from the World Bank into six pillars.

Ilm4tesis - AI tool integrated with GPT4 engine to create research questions from the thesis repository in the Department of Economics PUCP.

Working Papers

Impact of the Availability of Chat-GPT on Software Development: A Synthetic Difference-in-Differences Estimation using GitHub Data. Presented at The Munich Summer Institute and the Annual Meeting of the Academy of Management 2024

Can Market Competition Reduce Corruption? with Amen Jalal, Muhammad Haseeb, and Kate Vyborny. Presented at the 9th Zurich Conference on Public Finance in Developing Countries

High Dimensional Metrics in Julia with V. Chernozhukov, C. Hansen and M. Splinder.

Estimating Heterogeneous Effects of Cash Bonuses in Teacher Retention with Machine Learning: Evidence from Peru with S. Zhang and R. Tang.

Fertility and Education Patterns Across Different Phases of Development - Master Thesis.

Books

V. Chernozhukov and A. Quispe (2021). Inference on Causal and Structural Parameters using ML and AI with R, Python and Julia, used in the course 14.38 at MIT.

A. Quispe (2022). Machine Learning and Causal Inference using Python, used in the course MGTECON-634 at Stanford.

Awards, Fellowships, & Grants

2021	Award for Excellence in Teaching , PUCP	
2020	Emergency Grant , LMU Munich	€ 500
2019	PROSA-LMU Stipendium , DAAD to conduct research at Harvard University	€ 2,300
2017-2018	DAAD Kontakt Stipendium , German Academic Exchange Service	€ 1,400
2017	Student Exchange Scholarship , PUCP, to study in Germany	\$ 5,000
2012-2016	Full Scholarship for academic excellence in five consecutive years , PUCP	

Teaching Experience

2025	Advance Machine Learning I and II , Lecturer
2022	Causal Trees and Causal Forest Workshop using Python , Lecturer
2021-2025	Machine Learning and Causal Inference using R, Python and Julia , Lecturer
Fall 2018	Advanced Econometrics , Lecturer

INEI Peru
World Bank
PUCP-U. Pacífico
UNMSM

Skills

Programming: Python, Java, Julia, R, Stata-Mata, C++, Matlab, Git, Linux, HTML

Frameworks: PyTorch, TensorFlow, Apache Spark, MySQL, PostgreSQL, OPENAI-API, Azure, AWS

References

1. Prof. Victor Chernozhukov, Department of Economics & Center for Statistics, MIT, USA. Email: vchern@mit.edu, Phone: (617) 253-4767
1. Prof. Georgia Gkioxari, Computing + Mathematical Sciences (CMS) Department, Caltech, USA. Email: georgia.gkioxari@gmail.com, Phone: (626) 395-6580
2. Prof. Dietmar Harhoff. Director of the Max Planck Institute for Innovation and Competition. Email: dietmar.harhoff@ip.mpg.de, Phone: +49 89 24246-550.
3. Pedro Sant'Anna. Professor of Economics at Emory University Email: pedro.santanna@emory.edu, Phone: (404) 727-6364.