Leo Torres

Senior Data Engineer

leo@leotrs.com ◆ leotrs.com ◆ github.com/leotrs ◆ Leipzig/Berlin

Technical Skills

Languages & Frameworks

- Python (Expert): pandas, NumPy, SciPy, FastAPI, NetworkX, scikit-learn
- JavaScript (Advanced): Vue.js, HTMX, Node.js, D3.js
- AI & LLMs: Frontier LLMs (Claude, GPT, open models), RAG systems, vector databases, agentic workflows
- Other: SQL (advanced), Mathematica, C/C++, R, Julia

Infrastructure & Tools

- Cloud & Data: GCP (BigQuery, Dataflow, Cloud Run), Azure ML Studio
- Databases: BigQuery, PostgreSQL, experience with billion-row datasets
- DevOps: Docker, Airflow, GitHub Actions, Netlify, Fly.io
- Development: Git, pytest, CI/CD, OOP design, performance optimization

Experience

FGS Global — Lead Data Scientist → Tech Lead, Data Platform `May 2023 - Present`
Berlin, Germany

- Designed, built, and own end-to-end the news media analysis pipeline at FGS — an LLM-powered system processing 10M+ documents/day into client-facing intelligence for Fortune 500 firms
- Architected retrieval infrastructure over billion-row datasets in BigQuery with sub-second query performance
- Designed the RAG and vector-search layer that made domain-specific retrieval accurate enough for production client use
- Technical lead for a team of up to 6 engineers and data scientists; direct manager for up to 5, owning performance reviews; held hiring authority for select roles
- Built the FastAPI service layer serving multiple internal products from a shared data backend

Max Planck Institute - Postdoctoral Fellow `Aug 2021 - May 2023`
Leipzig, Germany

- Developed spectral graph theory algorithms for complex networks
- Co-led open-source software development for research community

Yahoo! Research - Research Intern `May 2019 - Jul 2019`
New York, NY

- Built graph learning models on terabyte-scale social network data
- Developed distributed computing pipelines using PySpark

Wolfram Research South America - Research Programmer `2012 - 2014`

Lima, Peru (Remote)

- Developed data pipelines for the Wolfram|Alpha knowledge engine
- Owned specific data domains end-to-end; worked in a globally distributed team

Open Source Projects

hns — Headless Multi-Agent Orchestration (private) `2025 - Present`

- Built a headless dispatcher that runs Claude Code workers in parallel through a state machine: PR triage, CI gating, stuck-detection with escalation, orphan recovery, autonomous worktree dispatch
- Python; designed the abstractions and control loop from first principles

The Aris Program — Founder & Lead Developer `2025 - Present`
`aris.pub`

- Founder of a multi-project open-source org building web-native infrastructure for academic publishing
- Designed RSM (Readable Science Markup), a markup language for scientific documents, with a tree-sitter grammar
- Built Scroll Press, a preprint server (FastAPI, PostgreSQL, HTMX) accepting Typst, Quarto, MyST, and Jupyter; live in beta
- Architecting a collaborative editor (Vue) as the reference RSM implementation

XGI (CompleX Group Interactions) - Co-Lead Developer `2021 - Present`
`xgi.readthedocs.io`

- Python library for hypergraph analysis, NumFOCUS affiliated
- `NumPy pandas pytest GitHub Actions`

Manim Community Edition - Core Developer `May 2020 - May 2021`
`manim.community`

- Mathematical animation library popularized by 3Blue1Brown
- Graphics programming, code review, community management

netrd - Co-Lead Developer `Jan 2019 - Jul 2019`
`github.com/netsiphd/netrd`

- Implemented 40+ algorithms for network reconstruction and comparison
- Co-led team of ~6 core developers; established coding standards
- Published in Journal of Open Source Software (JOSS)

Professional Development

- Talk: "LLMs: Computation or Measurement?" - Berlin Data Meetup, 2024 (`YouTube`)
- Workshop: "Communicating with Data" - Designed and delivered technical workshop for enterprise clients

- Recurse Center: 12-week programming retreat focused on algorithm design and software craftsmanship (2016)

Selected Publications

- L. Torres, A. S. Blevins, D. S. Bassett, T. Eliassi-Rad. "The Why, How, and When of Representations for Complex Systems." SIAM Review 63(3), 2021.
- J. Jost, R. Mulas, L. Torres. "Spectral theory of the non-backtracking Laplacian for graphs." Discrete Mathematics 346(10), 2023.
- Co-lead developer on XGI (JOSS, 2023) and netrd (JOSS, 2021); see Open Source Projects above. Full list on `Google Scholar`.

Education

Ph.D. Network Science - Northeastern University, Boston, MA `2016-2021`
Dissertation: Spectral Aspects of Mining Complex Networks
Advisor: Tina Eliassi-Rad

B.Sc. Mathematics - Pontificia Universidad Católica del Perú, Lima `2009-2015`
Top 3% grades in 75-year history of Sciences Department
Honors Thesis: Small-World Random Graphs Beyond Watts-Strogatz