

JOEL GRUS

joelgrus@gmail.com

twitter.com/joelgrus

github.com/joelgrus

Technical leader, software engineer, data scientist, contrarian, best-selling author, popular conference speaker, podcast host, stand-up comic, livecoder, angel investor, and strong generalist. Enthusiastic about well-designed software, readability, clean code, mentoring and learning, extracting value from data, AI / ML / NLP, and functional programming.

BOOKS

[Data Science from Scratch: First Principles with Python](#) (O'Reilly Media, 2nd ed. 2019)

[Thinking Spreadsheet](#) (self-published, 2010)

OPEN SOURCE CONTRIBUTIONS

[allenai/allennlp](#): core engineer for PyTorch-based deep-learning library for NLP research

[clab/dynet](#): created Scala bindings for C++ dynamic neural net library developed at CMU

WORK EXPERIENCE

[ALLEN INSTITUTE FOR ARTIFICIAL INTELLIGENCE](#) | Seattle, WA | 2016 - present

Senior Research Engineer

Research and engineering on the [AllenNLP](#) library. I do a little bit of everything: library design, code refactoring, adding features, devops, implementing and training models, building React demos, collaborating with researchers, mentoring, writing tutorials, supporting end-users, giving talks, fixing bugs, and all the rest.

Previously I worked on the [Aristo](#) project, teaching computers to answer science questions. My responsibilities spanned pure engineering tasks (e.g. building data pipelines, implementing RPC frameworks), pure research tasks (e.g. designing and building topic-specific question solvers, coming up with strategies to generate large numbers of science questions), and everything in between (e.g. creating Java wrappers and a Scala API for a C++ neural net library).

GOOGLE | Seattle, WA | 2014 - 2015

Software Engineer

Tech lead on Supply Chain Planning team. Responsible for technical design and implementation of backend services, gathering requirements, coordinating developers, writing code, and relationships with business users. Previously backend engineer on Display Ads; designed and built data-driven features for a greenfield ads benchmarking tool.

VOLOMETRIX | Seattle, WA | 2011 - 2014

Chief Scientist and Head of Analytics

Employee #2 at enterprise SaaS startup. Created prototype for in-memory analytical engine, designed and implemented most of the analytics in the product, and participated in every other aspect of growing the company from 3 people to 15 people. Responsibilities split roughly evenly among data science R&D, writing production code, and managing people. (Company acquired by Microsoft in 2015.)

DECIDE | Seattle, WA | 2010 - 2011

Analyst

Wrote code, built models, and designed processes to collect, analyze, and predict data for consumer electronics startup. (Company acquired by eBay in 2013.)

FARECAST / MICROSOFT | Seattle, WA | 2006 - 2010

Senior Analyst / "Fareologist" / Software Developer

Mined travel data for "fareology" insights, wrote code and built dashboards to track business metrics and customer activity, simulated revenue impact of changes to analytic products, designed algorithms to identify "deals". (Company acquired by Microsoft in 2008.)

TAHOMA CAPITAL | Bellevue, WA | 2006

Senior Analyst

Handled day-to-day operations of FX portfolio at hedge fund; modeled value and risk of proposed trades.

REPRESENTATIVE TALKS

- [Statistics Isn't That Funny, But It Has Its Moments](#), upcoming at SDSS 2019
- [Reproducibility: A Trojan Horse for Software Engineering Best Practices](#), upcoming at ICLR 2019
- [Modern NLP for Pre-Modern Practitioners](#), opening keynote at QCon.ai 2019
- [If Not Notebooks, Then What?](#) "Reproducible AI" workshop at AAAI 2019,
- [Writing Code for NLP Research](#) (with Matt Gardner + Mark Neumann), tutorial at EMNLP 2018
- [How Becoming Not a Data Scientist Made Me a Better Data Scientist](#), Southern Data Science 2018
- [AllenNLP and Best Practices for NLP Research](#), SAP/Concur "Lunch-n-Learn" (and others) 2018
- [I Don't Like Notebooks](#), JupyterCon 2018
- [Livecoding Madness: Let's Build a Deep Learning Library](#), ODSC West 2017 (also, OSCON 2018)
- [Using AI to Answer Science Questions](#), ODSC East 2017 (also, Southern Data Science 2017)
- [Fizz Buzz in Tensorflow](#), PyData Chicago 2016 (also, ODSC West 2016, WrangleConf 2016)
- [Fun With Trump Tweets](#), Seattle TwitterDev Meetup, October 2016
- [Learning and Teaching Data Science from First Principles](#), Data Day Texas 2016 (and others)
- [Stupid Itertools Tricks for Data Science](#), PyData Seattle 2015
- [Image Posterization with k-Means Clustering](#), Seattle PyLadies (workshop)
- [Secrets of Fire Truck Society](#), Ignite Lightning Talk at Strata 2013

EDUCATION

MS Social Science, Caltech

MS Mathematics, University of Washington

BA Mathematics, Rice University

FUN

[Adversarial Learning](#) (podcast co-hosted with Andrew Musselman)

[livecoding videos](#)

[Fizz Buzz in Tensorflow](#)

SELECTED PUBLICATIONS

- [Reasoning about Actions and State Changes by Injecting Commonsense Knowledge](#), EMNLP 2018.
- [AllenNLP: A Deep Semantic Natural Language Processing Platform](#), NLP OSS @ ACL 2018.