

Peter Demin

Senior Backend Engineer

peterdemin@gmail.com

+1 (240) 342-1438

<https://peter.demin.dev>

<https://linkedin.com/in/peterdemin/>

Skills & Tools

Languages: Python, Go, SQL, JavaScript, C++, Rust
Frameworks: Django, gRPC, aiohttp, Flask, React, Tailwind, 11ty
Libraries: Protobuf, Jinja2, NumPy/SciPy, Pandas, Llama2
Tools: AWS CloudFormation, Terraform, Git, Bazel, Kibana, Grafana, Jenkins

Experience

Senior AIML Engineer Apple, Seattle, WA Dec 2023 – Now

- Launched a new summarization model for Siri Web Answers using a fine-tuned 30B in-house LLM, improving the quality of answers for the end users and lowering the latency.
- Proactively built internal tooling for everyday operations, saving the team hours of routine tasks and speeding up onboarding for the new team members.

Senior Software Engineer Moveworks, Remote Nov 2019 – Sep 2023

- Provided technical leadership to several mission-critical long-term projects from requirements collection to support unlocking multiple enterprise deals with \$10M+ in ARR.
- Built continuous internationalization and localization pipelines for daily releases.
- Designed, built, and maintained Python microservices using gRPC and async I/O HTTP in Bazel-powered monorepo deployed to Kubernetes on AWS with Terraform.
- Onboarded, coached, and mentored engineers of different levels and experience.
- Proactively improved code quality, observability, knowledge sharing, engineer productivity, and adoption of best practices, cutting down on-call burden by several hours per week.

Senior Software Engineer National Institutes of Health, Bethesda, MD Oct 2015 – Oct 2019

- Designed and built several RESTful APIs with Django REST Framework to support the National Center of Biological Information.

Lead Software Engineer Epam, Minsk, Belarus Apr 2015 – Sep 2015

- Developed cluster configuration management tools on top of SaltStack that allowed deploying and upgrading air-gapped clusters of storage servers with Apache Hadoop family, GPFS, SNMP, and proprietary software.

Software Engineer Yandex, Moscow, Russia Oct 2013 – Apr 2015

- Developed several RESTful API services for high-load mobile-based applications using SQL and NoSQL DBs.

Education

- Scrum fundamentals and Agile planning techniques 2015
- Machine Learning (Andrew Ng) Coursera 2014
- Moscow Aviation Institute Masters in Robotics and Computer Science 2005 – 2011