Connor Brinton

connor@brintonium.com
(571) 271-4574
LinkedIn
GitHub

I'm a machine learning engineer with extensive experience building, scaling and iteratively improving high-performance deep learning models, ranging from fine-tuned NER models to highly optimized self-hosted LLMs.

Technologies

Programming and Configuration Languages	Python, Cython, Go, Typescript, JavaScript, Bash, Zsh, SQL, Dockerfiles, YAML, TOML, JSON, HCL (Terraform), Starlark, C++, Java, R, etc.
Machine Learning Frameworks	HuggingFace Transformers, Scikit-Learn, TensorFlow, PyTorch, PyTorch Lightning, SpaCy, Thinc, Flair, NLTK
Machine Learning Libraries	NumPy, Pandas, SciPy, Seaborn, Matplotlib, Plotly, Bokeh, Ray Core, Optuna, TensorBoard, Gradio, HuggingFace Datasets, HuggingFace Tokenizers, NetworkX, Joblib
Web Frameworks & APIs	FastAPI, Starlette, Uvicorn, Tornado, Sanic, Flask, BentoML, gRPC, BetterProto, HTTPX, Requests, GraphQL, Ariadne
Cloud & Infrastructure	AWS, GCP, Kubernetes, Terraform, Pulumi, Docker, BuildKit, Argo Workflows, Datadog, PostgreSQL, Snowflake, CircleCI, GitHub Actions, Cirrus CI, Git
Python Ecosystem	Poetry, uv, pyenv, PyTest, Mypy, Ruff, Black, Pylint, Pydantic, Typer, Rich, SQLAlchemy, loguru, tqdm

Recent Experience

Nightfall AI

November 2023 - Present — Data Science Team Manager
December 2022 - November 2023 — Senior Machine Learning Engineer

- Optimized a Transformer-based inference service, **tripling throughput and reducing latency 21x** by migrating from BentoML to Ray Core, introducing an in-memory cache, quantizing the model and leveraging FlashAttention
- Improved training speed of a **260M** parameter Vision-Language Model (VLM) **4x** by rearchitecting the data loading process, fixing bugs preventing the use of multiple GPUs and caching prompt tokens properly
- Improved precision of 10 sensitive information detectors by **30%** by training a convolutional neural network (CNN) model on linguistic token features to filter predictions from a fast entity detection pipeline
- Standardized our company-wide annotation data format to minimize model-specific data loading logic and developed team tools for analysis, validation, repair, prediction, evaluation and visualization of our data and model predictions
- Identified and managed high-impact initiatives, guiding a team of 4 data scientists to success, including:
  - Reduced false discovery rate (FDR) **2.5x** (**46%** → **17%**) and increased recall **1.5x** (**42%** → **69%**) across our sensitive information detectors by improving data quality, model architectures and hyperparameter tuning
  - Migrating to a new OCR model with **30%** better character recognition accuracy
  - Expanded international support to European, South American and Southeast Asian locales
  - Rapidly responding to client feedback with one-click training of all of our production models

More experience on next page

Zoom February 2021 - December 2022 Senior Machine Learning Engineer
Solvvy (acquired by Zoom) June 2018 - January 2021 Data Scientist

- Led data science team strategy and day-to-day operations, including roadmap planning, hiring, team assignments, daily standups, weekly team meetings and cross-department communication
- Increased end-user self-service rates from **27%** to **54%** by enhancing our deep learning-based question-answering and intent detection models, resulting in a **33%** reduction in client support costs
- Pioneered use of fine-tuned multilingual foundation models to expand and enhance language support and cross-lingual transfer capabilities across our machine learning products
- Maintained extremely high uptime of our machine learning services and managed growth from less than **1M** to nearly **7M** sessions per month, by continually enhancing our Kubernetes-based machine learning platform
- Significantly reduced code review and new team member onboarding time by introducing a standardized set of CI checks for Python repositories, including Mypy, Pylint and Black

Mosaic ATM Summer 2017 Machine Learning Researcher

- Refined and implemented an algorithm for interpretable decomposition of arbitrary datasets along meaningful basis vectors as part of contract work for DARPA
- Developed an expectation-maximization-based inventory prediction algorithm for a large retailer
- Participated in daily calls with external customers, explaining recent progress, identifying immediate challenges and recommending solutions

Dror Lab at Stanford University Summer 2016 Computational Researcher

- Resolved long-standing challenges in molecular dynamics simulation by writing automated software for simulation maintenance, which achieved widespread use throughout the lab
- Performed over 100 simulations of complex protein-protein interaction systems over long time periods (~4 μs each) using Stanford's high performance computing resources and modern simulation software
- Analyzed geometric differences between each simulation to produce novel hypotheses regarding functional protein differences, and presented findings to collaborating labs

Open-Source Contributions

- **TensorFlow:** Loosen version constraints on `typing-extensions`
- **TensorFlow Text:** Fix macOS packaging
- **SpaCy:** Add Mypy check to CI and introduce many type annotation improvements
- **SpaCy:** Escape annotated HTML tags in span renderer
- **Thinc:** Fix Mypy plugin crash on variadic arguments
- **Flair:** Show allennlp import errors when installed
- **Poetry:** Fix request authentication when credentials are included in URLs
- **Typeshed:** Fix incorrectly named parameters in logging module
- **Popeye:** Add arm64 Darwin releases to krew-index
- **GQL:** Add type overloads for get_execution_result
- **Graphene-SQLAlchemy:** Don't suppress SQLAlchemy errors when mapping classes
- Documentation contributions to HF Transformers, SpaCy, Boltons, Ariadne, and Cirrus CI

Education

Stanford University	Master of Science (M.S.) in Biomedical Informatics	2019
Stanford University	Bachelor of Science (B.S.) in Computer Science *Emphasis in Artificial Intelligence (AI)*	2018

In my free time I enjoy learning new board games, reading Brandon Sanderson books and playing around with home automation software. I'd love to chat about any of them with you!