Brian Fitzgerald

PERSONAL DETAILS

Location	Columbia, MO / Remote
Phone	(618) 616-3533
Website	`https://brianfitzgerald.xyz/`
Mail	`brianfitzgerald242@gmail.com`
GitHub	`https://github.com/brianfitzgerald`

WORK EXPERIENCE

ML Research Engineer 2023 -
Stability AI

- Co-led the post-training for Stable Diffusion 3, which involved fine-tuning the model to improve aesthetic quality and prompt alignment, as well as determining the final look and feel of the model's output.
- Led the safety training for Stable Diffusion 3, which involved developing a set of safety metrics and tuning the model to reduce the likelihood of generating unsafe or inappropriate content.
- Trained a sequence-to-sequence model to perform prompt upsampling for Stable Diffusion 3, achieving performance in model size and inference speed that beats the state-of-the-art for this use case, while ensuring that the resulting image outputs are inclusive, culturally appropriate, safe, and high-quality.
- Worked on a variety of customer-facing LLM projects related to tool use and conversational AI, including fine-tuning, evaluation, and inference, across a variety of foundation language models. Worked with researchers and engineers across the company to achieve best-in-class safety and accuracy scores for each use case.
- Trained, deployed, and maintained a text classifier for user prompt moderation, and an image classifier for generated content moderation. These classifiers meet industry standards for accuracy on all safety-related categories tested, as well as additional ones specific to generative models.
- Worked on the Stability Fine-Tuning API. Developed an inference, fine-tuning, and post-processing stack that led to best-in-class fine-tuning training times, and best-in-industry output quality. This API is being trialed with several large enterprise customers.

Machine Learning Engineer, Inference 2022 - 2023
Stability AI

- Co-designed and wrote Stability's LLM inference solution, including custom kernels that add support for FlashAttention 2 to Stability's StableLM series of models, as well as other optimizations.
- Worked on Stability's next-generation text-to-image inference platform, as well as model compilation strategies to bring inference times and memory usage in line with industry standards, across a range of accelerators, including TPUs, Gaudi, Inferentia, and Nvidia GPUs.
- Developed prototype integrations for Blender and Photoshop to demonstrate uses of Stable Diffusion and LLMs within creative workflows. Shipped and maintained the Blender integration, which surpassed 50k downloads within 24 hours of its release.

Senior Software Engineer, Robotics 2020 - 2022
EquipmentShare

- Worked on various computer vision tasks for an autonomous construction system, including filtering and segmentation of LIDAR and image data in C++ and Python.
- Developed and maintained a distributed simulation environment for testing and developing autonomous construction equipment, including a highly performant LIDAR simulator and test scenario generator with Mujoco and Unity.
- Wrote and implemented a system for training large-scale object detection models, using training and evaluation data gathered from real-world testing.
- Developed a pipeline for the recording and replay of sensor data from physical machines, including a real-time visualization system for reviewing defects and anomalies in training data.
- Designed and implemented a custom video streaming solution intended for low-bandwidth use, streaming video from the robot's cameras to a remote client, as well as other services on the machine.
- Mentored a team of interns in developing a custom depth estimation and segmentation system for obstacle detection from video data.

Senior Software Engineer, Elogs 2019 - 2020
EquipmentShare

- Worked on expanding support for the Elogs ELD compliance system, which gave real-time alerts for FMCSA guideline violations to thousands of drivers. Grew the list of supported state regulations from 2 to 30.
- Maintained a real-time rule engine for processing and alerting on driver logs, which processed millions of logs daily from tens of thousands of active users.

- Maintained the React Native - based iOS and Android Elogs app, which served a wide range of mobile devices and trackers.
- Coordinated with other teams within the company to expand DOT regulation support across multiple products.

Software Engineer 2016 - 2019
CARFAX

- Led the development of the CARFAX Service Shops application, which provides shops with a toolset for managing customer-facing data for their shop and viewing analytics.
- Designed and implemented a package-based frontend architecture for myCARFAX's core products over the span of several project timelines. Developed a standard for sharing common functionality across products and teams.
- Created an architecture for one of CARFAX's first user-facing products running on Amazon Web Services. Collaborated on the strategy for migrating the CARFAX's on-premises consumer services to an AWS-managed, Kubernetes-based infrastructure.
- Worked on the rewrite of myCARFAX's core web applications from Angular to React/Redux. Worked with the User Experience team to enforce a consistent look and feel across myCARFAX products.
- Led the development of the CARFAX Service Shops application, written in React / Node / Typescript, which is used by 1,000s of customers daily.

Developer February 2018 - January 2019
InteraXon Inc

- Developed the CES demo experience for the MUSE 2 by Interaxon, an AR experience that allows users to try out a virtual version of the headset and explore its features. Built in Unity with ARKit, utilizing face detection and masking.
- Was later deployed in demo kiosks in California, and integrated into InteraXon's marketing material.

Developer September 2018 - February 2019
Healium XR

- Designed and developed several AR experiences for assisting with meditation in Unity with ARKit and ARCore, targeting mobile devices and the Oculus Go headset.
- Built a dashboard and analytics platform for tracking biometric data. Used AWS Cognito for authentication and Lambda and DynamoDB for the backend.

Developer June 2017 - May 2018
GeneTrait Laboratories

- Developed the MedTrait reporting system for GeneTrait.
- Developed a backend service to generate genetic testing reports with Node / Express, as well as a component library in React and Typescript to render the reports.
- Was involved in the entire product development cycle, including designing features and developing a user experience and visual identity.

Developer Oct 2015 - Jun 2016
QuarkWorks

- Worked on a variety of native mobile applications, including ZephyrCharts, an aviation mapping application for pilots. This involved developing native iOS and Android clients for each project.
- Developed the companion app for Columbia's local Roots n Blues festival, as well as a backend for managing user data in Google Cloud.

Software Developer Feb 2014 - Sep 2015
WireCloud, LLC

- Worked on various contracting projects including the backend for a local food delivery service. Built a system for managing a stream of orders, and assigning them to drivers, as well as a customer-facing order tracking system.
- Developed a companion mobile web application for drivers.

SKILLS AND TECHNOLOGIES

ML	PyTorch, Jax, Lightning, CUDA, Slurm, Ray
Vision	PCL, OpenCV, OpenGL, PhysX, Mujoco
Robotics	C++, Rust, ZeroMQ, Protobuf, ROS, Waymo Honeycomb
Frontend	TypeScript, React, Redux, MobX
Backend	Golang, Terraform, Kubernetes
AWS	SQS, CloudFront / S3, ECS, Lambda, Cognito, Cloudflare Workers
Mobile	Swift / iOS, Android SDK, SceneKit
XR	Unity, C#, ShaderLab, ARKit / ARCore, SteamVR SDK

EDUCATION

Governor French Academy 2011 - 2015
High School

University of Missouri 2016 - 2017
College, Not Completed