

# Levi DeHaan

[levi@levidehaan.com](mailto:levi@levidehaan.com)

<https://www.levidehaan.com/>

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

## TECHNICAL SKILLS

- **AI & Automation:** Large Language Model fine-tuning, ML, AI serving infra, llama.cpp, custom harnesses like langchain, advanced rag systems, systems management agents, ai trading intelligence, ai hardware solutions. Experience using and building automations using open and closed models.
- **DevOps & Cloud:** Kubernetes, Docker, Jenkins, Terraform, AWS, Azure, GCP, Databricks, terraform, playwright,
- **Programming:** JavaScript/TypeScript, Golang, C++, Python, Swift, Java, C#, PHP
- **Web, Desktop & Mobile:** Next.js, React, Node.js, SwiftUI, AppKit, Chrome Native Messaging, Cordova
- **Databases:** MySQL, Cassandra, Kafka, Spark, PostgreSQL, SQLite (GRDB), Apache Arrow
- **Virtualization & Networking:** Xen, VMWare, Cisco, HAProxy, WebSockets, gRPC
- **Extended Reality (XR):** Virtual Reality, Augmented Reality, Unity 3D
- **System Administration:** Linux, Windows, Server Administration, Datacenter management
- **Other:** Framework Design, Project Management, custom ai solutions

## EXPERIENCE

### **Glean | AI Outcomes Manager & Forward Deployed Engineer**

Nov 2025 – Present

At Glean, I enable clients to maximize AI value by building agents, complex integrations, and hybrid AI ecosystems.

I manage three times the average client load while significantly exceeding quarterly targets.

Leveraging deep experience in software stacks, automation, and infrastructure, I provide diverse companies with critical technical guidance and implementation support.

I have also built many internal software applications that leverage glean:

- **QuestionMonitor:** Architected a production-ready, cross-process Swift and Python macOS application bridging a SwiftUI frontend to a Python PyTorch backend via a custom localhost WebSocket server. Engineered a sub-250ms latency ML pipeline utilizing Silero VAD (reducing hallucinations by 90-95%), faster-whisper, and custom-trained SetFit/transformer models.
- **VideoEditorPro:** Developed a local-first video editing application leveraging React, Node.js, and

FFmpeg pipelines for pixel-perfect canvas encoding, compositing, and text-to-speech (TTS) generation.

- **GTMExtensions:** Engineered an end-to-end training pipeline utilizing Chrome Native Messaging, TypeScript 5.x, and Webpack 5 to bridge browser extensions securely to macOS desktop applications via local TCP sockets.
- **Meeting Prep:** Created a macOS menu bar application using EventKit for Apple Calendar sync and GRDB/SQLite for structured, durable AI agent meeting preparation.
- **Agentic Workflows:** Designed complex, multi-step agent architectures with advanced classification logic with my clients.
- **Core Infrastructure Debugging:** Bridged the gap between customer deployments and core engineering by debugging complex AI infrastructure issues at the code level, investigating sandbox-orchestrator crashes, and designing highly secure, multi-VPC AWS Transit Gateway architectures for enterprise MCP server connectivity.

## **AutomatonAutomation.ai & DeHaan Consulting | Principal Architect & Consultant**

Dec 2018 – Present

- **AI Cluster:** Architected a high-performance go server utilizing llama.cpp that natively integrates MCP servers for autonomous tool selection, managing high-speed agent context memory, agents, assistant, omni voice chat, and many other tools.
- **BifrostPipe:** Engineered a highly performant C++ Kafka-centric data pipeline handling real-time high-frequency streaming via WebSockets and kafka for low-latency market data transformation.
- **Alpaca Auto Trader (AAT):** AlpacaAutoTrader is an autonomous AI-powered stock trading system that uses a cooperative two-stage pipeline: a fast 7B scanner model identifies candidates and executes trades. The system runs 24/7 with a comprehensive intelligence stack including automated news/research streaming, Twitter signal analysis, QMD semantic search across 100K+ documents, TimesFM price forecasting, and automated learning pipelines that generate trade autopsies and adjust signal weights.
- **FluentD LLM Server:** Built an end-to-end log analysis solution combining FluentD log aggregation with custom-trained Microsoft Phi-2 models running on Kubernetes for real-time security threat detection and automated anomaly response.
- **Visual & Video Pipelines:** Created the Kafka Visual Data Pipeline for building stream processing through an intuitive drag-and-drop interface, and the YouTube Financial Video Analyzer which extracts market intelligence from multimodal video content using DeepSeek analysis.

## **Petco - Online Retail**

Automation and AI Systems Architect 10/2024 – 03/2025

In my tenure as Automation and Artificial Intelligence Architect at Petco, I spearheaded the development of a cutting-edge Large Language Model serving framework that transformed the company's technological capabilities. I designed and deployed a dynamic AI system across AWS, Azure, and an in-house GPU cluster, demonstrating my expertise in managing diverse platforms and establishing a scalable foundation for Petco's AI initiatives.

My efforts delivered standardized, repeatable workflows that simplified AI adoption and ensured operational consistency. By architecting a flexible, multi-platform infrastructure, I reduced costs and unlocked tremendous potential for growth, positioning Petco to thrive in an AI-centric future.

- **LLM:** Built advanced LLM systems using open source technologies like llama.cpp and transformers for incredibly fast inference on low cost systems.
- **AI Everywhere:** Deployed on multiple clouds, with Kubernetes backends, utilized K8's api for distributing inference. Can run on CPU and GPU using all resources. Memory and model management interfaces through Simple Ai Router.
- **Enhanced Applications:** Developed and deployed a cost-effective AI based solution to replace a google ads service, leveraging AI workflows and Simple AI backend service in a week.
- **AI Assist:** Mastered AI-assisted coding tools like Windsurf, Cursor, and Claude Desktop, establishing usage methodologies for effective and fast development.
- **Consistent:** Created standardized workflows for AI integration across development teams.
- **Scalable Systems:** Designed and deployed scalable AI infrastructure, reducing costs while enhancing capabilities.

## **Petco - Online Retail**

Manager - DevOps COE/SRE/Cloud Infrastructure Services + AI Systems Architect 08/2022 – 10/2024

In my tenure as a Manager at Petco, I've had the privilege of spearheading multiple teams and initiatives that have significantly enhanced our operational efficiency and technological capabilities. I've been able to accomplish this by leveraging my experience to build a team of highly skilled engineers and architects. I regularly assist in interviewing and evaluating candidates for other teams. I was able to take one of my teams from a 6 month ticket closure time to hours by automating systems with ai and reallocating resources efficiently.

- **Rapid Advancement:** Within a span of six months, I was entrusted with the leadership of the DevOps, SRE, and Cloud Infrastructure Services (CIS) teams, reflecting my capability to drive results and manage diverse teams.
- **Developer Engagement:** Initiated an internal advertising campaign within the first two months, educating hundreds of developers about the capabilities and tools of the DevOps team. This resulted in an increased collaboration across teams.
- **AI Everywhere:** My team(s) spearheaded the development of several AI based tools: Ticket Whisperer, PetChat, and many more teams use our team for AI to build their own apps.
- **Operational Efficiency:** Centralized our logging and statistics into a unified dashboard, streamlining the SRE's collaboration with the DevOps team and ensuring data-driven solutions.
- **AI Integration:** We deployed an Azure OpenAI Fine-Tuned Curie and then Davinci model, automating ticket responses for administrative tasks. This innovation liberated resources, allowing us to focus on more strategic initiatives.
- **Cloud Deployment:** Successfully implemented a custom Backstage system, integrating it with our Terraform tools to facilitate seamless deployments across various cloud environments. This system is also integrated with our internal user management for efficient authorization.
- **AI Chat Framework:** Designed and launched an advanced AI chatbot framework, powered by a

customized Langchain agent tying into confluence. This framework provides instant answers to complex, context-specific queries, transforming hours of research into minutes.

- **Model Fine-Tuning:** Established a robust pipeline to fine-tune models on azure openai for our Ticket Whisperer system.
- **Pet-Centric Solutions:** Pioneered the development of AI agents tailored for pet-related tools, enhancing both customer experience and providing veterinarians with contextual information instantly.
- **Team Transformation:** Before my leadership, a team I was given to manage had a ticket closure time of up to six months. Within two months, I revolutionized the process, reducing response times to mere hours and automating numerous systems. This transformation led to a significant reduction in daily tickets and allowed us to reallocate resources more efficiently.
- **Collaborative Leadership:** As my role evolved into a more strategic one, I ensured smooth operations by onboarding talent to manage day-to-day reporting. This allowed me to focus on crafting solutions and collaborating with other teams to kickstart projects.

## **Healthpointe Solutions - AI Healthcare**

Director - Cloud Infrastructure 03/2021 – 08/2022

- Designed and built the base infrastructure for the business using Terraform on AWS/Azure when I worked at Enable Data. I was hired as a jack of all trades to get HPS up and running quickly and securely. This was a continuation of my work at Enable Data
- Designed and built Terraform Manager, an internal project for managing hundreds of different terraform configurations allowing easy push button deployment of multiple configurations to any cloud.
- Designed and built Imaging Manager using Packer and Ansible, an internal project for managing the automation of machine images for use on all platforms with built in security checks, virus checks, and testing before pushing new images to dev/test/prod environments for further testing.
- Designed and built TerraMatch - a system for ensuring that the only systems running are the ones deployed by Terraform in order to ensure systems are exactly to spec.
- Designed and built the entire process for deploying Databricks on our Azure and AWS private cloud and the process for ingesting/transforming client data. I wrote the software for processing and analyzing Patient data from healthcare/insurance companies on Databricks using a mix of python and scala. I built and managed the custom docker images for our Databricks clusters and worked with our Data Science teams to better understand their needs and implement tools in those images to make their jobs easier.
- Worked on a Successful POC with Dell ECS where I designed and deployed a Kubernetes + Tableau based solution so Dell can offer HPS products on their ECS platform to healthcare providers.
- Worked on HITRUST certification security implementations, and procedures. Security Security. C Suite security/network consultant for international travel to our potential clients.
- Mentored and Educated fellow employees on any technical matters they were interested in or needed to know to be successful at their jobs.

## **Enable Data Consulting**

#### Sr. Systems Architect 10/2018 – 03/2021

- Designed and built a scalable web application for Greenwich.HR which scans the internet for changes in specific web based properties in order to provide Labor Market Intelligence. The application launches and manages thousands of ECS jobs and tracks their progress. The system has built in job creation, reporting, customer management, and a fully manageable queue system.
- Converted a production R project to the Databricks spark platform for IRI Worldwide as a poc to show the possibility of converting all their R systems to Databricks. I was able to get the project to run in 15% of the production run time; saving time and money as well as reducing complexity of the project. I also tested out other solutions such as building out a test Kubernetes cluster running spark jobs managed using Jupyter.
- Worked with a team to build the base infrastructure, security, processes and automation infrastructure for HealthPointe Solutions Cognitive AI. Using terraform, ansible, and jenkins we built a system to automatically deploy the entire infrastructure on multiple platforms like aws and azure. We defined the security posture and built the security into the system using automated tooling to scan, log and detect every aspect of the systems in order to provide the best security solution possible.
- Initial Interviews for client companies ranging from low level spark to high level systems management positions, and everything between.

#### **Nielsen ( Contracted through Oxford )**

##### Sr. DevOps Architect specializing in Kubernetes 02/2018 – 9/2018

- Designed and built Azure AKS/ACS interface to launch big data jobs into Kubernetes clusters using Jenkins
- Created a new and faster workflow to save time on job runs and on the time it takes to create them.
- Brought best practices and upgraded the team with CI/CD practices for faster code delivery and tracking.
- Trained other departments in the company on how to deploy jobs as quickly as ours using in person training and documentation.
- Worked on documenting that allowed us to bring our new team mates up to speed much faster.
- Designed and built a GUI interface for the kubernetes/jenkins system allowing the creation and deletion of clusters/jenkins installs/jobs ( Using Vue.js / Express ).
- Reduced big data deployment time and improved big data workflow.
- Helped my team improve their scripting and backend skill sets as well as educating them on containerization and orchestration systems like kubernetes and mesos.

#### **De Haan Consulting, LTD**

##### Sr. Systems Administrator / Sr. DevOps Manager / Virtual Reality / Augmented Reality Oct 2012 – Current

##### **Internal:**

- Designed and developed software platform for managing server systems in virtual reality named

ORCA. <https://youtu.be/zTdES-rOgcY> (VR Video)

- Designed Microsoft Hololens application for making oil refinery inspections easier to verify and manage.

## **Recent Clients:**

### **VR Garage** - Jan 2017-Feb 2018

RG is building a toolset that enables users to easily create VR video / image / 3D model experiences using our web based platform.

I designed the first Unity VR Concept at the beginning, and a team of developers took it to its conclusion. While they were working on that I spearheaded the web based experience. My team and I built it out using JHipster microservices and A-Frame.

I designed the backend services of the platform around fabric8 running on kubernetes. It's a very performant system and has enabled us to easily add to our engineering team with very little ramp up time (our newest guy was fully setup with full ide integration with all the projects and cluster within a day).

One of the huge benefits of microservices and clustering is the ability to dynamically scale services up and down to keep the cluster balanced and requiring far fewer machines, the cluster is only costing a 120 a month on linode servers, but is running a full toolstack and production ready services with blue green jenkins pipelines, full integration between services, such as a image processing service that converts all 360 images between mono and stereo, resizes based on bandwidth, etc..

I was able to design everything from the ground up, which is a rare and fun occurrence.. **Simple Energy** - 2017-2017

- Trained a local company to assume my full time tasks.
- Reduced future deploy complexity through scripting for common tasks.
- Assisted in transition to aws, provided insight into best ways to migrate existing infrastructure.

## **Simple Energy**

DevOps Manager 08/2015 – 04/2017

- Created a new DevOps workflow focused on improving automation and system deployments. • Took deploy time from 3-4 weeks to less than a day, and platform updates to minutes from hours. • Instantiated the use of Docker/Mesos to reduce deploy times and increase uptime on volatile systems.
- Instantiated the use of Jenkins/GitLab to increase developer productivity and accountability. • Reduced time to test for developers and reduced complexity to get to production. • Improved testing strategies and implemented automated recovery for simple systems. • Continuing to reduce costs across the board by further automating services and deployments to more cost effective methodologies (such as GCP/AWS) when possible.
- Worked on and deployed Various big data systems working with Scala and Python to help clients better understand energy usage.

## **Promet Source**

Sr. Developer / Sr. Systems Administrator / DevOps Manager 01/2015 – 07/2015 • Helped design and start new implementation and build of the DevOps team. • Transferred important servers from Promet's various clients onto a similar monitoring/deployment

and server infrastructure to allow for smoother deploys and to reduce time to production for developers.

- Designed and built a new server infrastructure using Google Cloud, saving money for the company and our clients. Some clients saved more than half what they used to pay moving to the new infrastructure.
- Worked within the infrastructure to implement new tool stacks to increase the efficiency of the team and to allow the DevOps team to incorporate automation in their processes. We successfully went from several days to do a deployment to several hours to get the same amount of work done.
- Managed several development projects utilizing multiple programming languages, php, javascript, java, and python.
  - Worked on a new drupal module that ties in Tika and Solr to index millions of pdf's efficiently on a small server using an incremental job runner that can run the downloading, parsing and indexing jobs during low server usage.
  - Built a Jenkins automation cluster to run all various types of jobs from backups to builds and the team is adding new things daily as they learn how to better utilize the tool. ■ Designed and implemented an Elasticsearch cluster service using Docker, Google Cloud, and Kubernetes that stores the logs from all sites and servers enable more efficient troubleshooting and debugging.
  - Developed a node.js application using Tika to convert pdf's to text and made the data available via a restful api. It has full caching and cache control as well.
  - Using the data from elasticsearch in conjunction with a monitoring framework that notifies the team when a service goes down or passes a load threshold.
  - Developed an IMAP mail application using node.js and express for the server and Angular.js for the frontend UI.

## **Apex Systems/Time Warner**

Systems Administrator/DevOps Engineer 05/2014 – 11/2014

- Utilizing tools like Jenkins, Docker, Mesos, and Puppet my team and I designed and implement new processes for continuous integration to enable the data scientists on our team to quickly write and test their code, allowing for greater efficiency, compartmentalization of their builds, minimized debug time, better enabled dry practices and a greater focus on the code not the systems.
- I built new processes for onboarding using a tool called Vagrant to build customized deploys that the new hires could load onto their machines to become productive much faster than setting them up on the development servers and all the other endless tasks that used to be done to onboard someone.
- Introduced Docker as a tool for development and system administration:
  - Designed big-data workflows using docker to containerize the platforms that we were using

on bare metal. This saved a lot of money on hardware because we could use one of our big servers for more than just a couple processes. This enabled us to create clusters within a single machine to test out larger deployments using the same stack.

- Designed an entire big-data workflow around mesos/marathon and built a UI in angular.js to accompany the backend technology in order to utilize consul and registrator.
- The developers are able to pull the most recent image from any deployed server and instantly have it up and running. This enables developers to quickly iterate on code changes, share those changes, and it provides a reliable testing environment because everyone has access to the same images.

## **De Haan Consulting, LTD**

Team Lead / Software Architect / Sr. Programmer / Sr. Systems Administrator / System Architect  
10/2012 – Present

### **Feb - April 2013**

I worked with a company based out of Omaha as a JavaScript expert consultant. The job was rescuing a failing project before it was shut down.

The project was a Java platform serving a javascript frontend. One of the core problems were the programmers who initially wrote the first iterations of the product. They were writing javascript like Java code. After a couple sessions discussing the current code structure, its issues, and what we could do about them we created policies to streamline the process, and a timeline to fix the bugs and submitted them to the executives for approval. They did not want to move forward with the changes and after transitioning out, the project was shut down due to system failure.

### **Jan – Feb 2013**

I worked with a company called WeatherCloud to develop an android application that connects to several bluetooth transmitters affixed to cars and converts and processes that data within a second to show real time stats to the driver.

The application was written using apache cordova and angular.js for the frontend. I wrote a bluetooth LE plugin for cordova to allow for the fastest transfer of data from the devices without slowing the phone down too much. The first iteration of the project was a success and I delivered an alpha to them to show to investors.

### **Nov - Dec 2013**

I worked with a company called o2o developing a yellow pages type application for argentina. It was written in php using a customized framework developed by o2o. The task I was given was to introduce a better onboarding process for new programmers, to integrate Docker into the development workflow, and the rest of my duties were to write customized modules for the framework and the frontend in javascript.

I was able to design and build an onboarding platform using Docker and Vagrant to get new programmers ready to code immediately. Allowing them time to go and change things in the framework without breaking our development servers copy.

The introduction of Docker was very helpful in providing a clean image that could be pushed to testing, and then production without the hassle of setting up all the needed services and specific

versions of packages needed by the framework.

## **Oct - May 2012**

I worked with a company called Pearson Education developing a customized javascript framework so they could easily write frontend applications for their student portal. Initially I started out writing it myself, as the project progressed my team grew, and we implemented agile methodologies to improve our workflow. We had to design and implement an entire stack from the backend server architecture using node.js to the front end framework which the company named Pearson.js. We also had to provide documentation and training to the individuals who would be taking our place and so we designed our framework to be as user friendly as possible by extrapolating the complexities and taking care of them in our framework.

## **appendTo, LLC**

Sr. Programmer / Systems Administrator 07/2010 – 09/2012

- Managed the servers for the business and clients, securing, updating, and configuring them for speed and reliability.
- Designed and developed server architectures for clients using many different platforms, frameworks and languages. Most commonly php based frameworks and javascript based platforms.
- Helped create and steer the culture towards a fun and supportive environment, and did interviews with potential hires. It was challenging to do this because we were a fully remote company. We primarily used Skype for communication.
- Created the onboarding process with my team in order to get our new hires up to speed quickly. Which was challenging because we had no standard systems and so I had to build for all major operating systems. We used virtualbox and some custom images to build out environments to get people up to speed quickly.
- Developed a chat robot that watched Skype chat rooms and listened for questions and also would give you stats on servers and code bases when asked. It was our version of hipchat.
- Spearheaded the website design and mobile development from the ground up for many projects including some internal projects that helped manage servers autonomously.
- Helped create and nurture relationships with clients like research in motion and Microsoft.

## **Springs Hosting**

Programmer / Systems Administrator 01/2010 – 07/2010

- Wrote custom applications in several different programming languages as well as implementing custom solutions/modules for different CMS's such as Drupal, ELGG, Joomla, wordpress etc.
- Designed and implemented the current VPS hosting solution using XEN and HyperVM, and customized HyperVM's core to implement certain aspects of XEN that were missing from the HyperVM Platform at the time.
- Applied several Aegir Distributions for clients and internal use. Tied Aegir into VPS distributions with custom packages tailored to a client's specific Drupal needs.
- Managed several clients who use the VPS platform to design and keep their systems running at optimum efficiency.

## **Monarch Digital**

Application Developer/Sys Admin/Web 2.0 Programmer 04/2009 – 11/2009

- Built drupal websites, wrote custom modules, rewrote backends, managed customer servers, managed monarch servers, designed client server architecture, created backup solutions that worked, internal tech support, client web site/server tech support.
- Defined the standards for coding, quality, and server builds.
- Trained new employees on php coding standards and built vm's for them so they could get into a project and start working right away.
- Defined and implemented business ideas, identified potential clients and earned their confidence.

## **Keystone Symposia**

Web Designer / Systems Analyst 04/2008 – 11/2008

I worked on the web platform which was written in cfml, at the time a new language for me. I caught on quickly and started closing tickets. My primary job was to assist the senior programmers with basic tasks until I learned enough to start working on more complex systems.

## **Rof.net**

Programmer 03/2006 – 03/2007

I was hired to do customer tech support and wireless installations. We got so many calls about web design that I finally convinced my boss to let me start building websites. I'd already done a few for CMC using php and so I went from support to web development. My first client's website is still up and running and I built it using one of the earlier versions of drupal.

## **DeHaan Design and Development**

Application Developer/Systems Administrator Consultant 01/1996 – 05/2010

We focused on our customers' needs and delivered a product that goes above and beyond their expectations. We understood that money is hard to come by in this recession. We created many solutions for our customers to help ease the financial strains. We didn't stop at just providing a site, application or service. We also wanted to help our customers build their customer base by creating solutions that helped them attract business and meet their customer's needs.

## **EDUCATION**

### **Colorado Mountain College**

2004 - 2007

I took classes for programming, writing, and math. I was thinking about getting a degree but ended up starting a company writing software for local nonprofits and local businesses.

### **High School**

1996-2001

I started attending college classes at 13, mostly programming and networking classes. I was hired as

a system administrator for all the campuses when I was 15 after taking the cisco networking academy and becoming a TA.

### **ASSOCIATIONS**

- Ran a javascript/node.js meetup when I lived in Colorado Springs.
- I started the Boulder Google Developers Group, we meet a couple times a month at the google campus in Boulder.