

# JOHN ZOLTON

Patent Attorney turned ML Engineer | [john@zolton.xyz](mailto:john@zolton.xyz) | [github.com/JohnZolton](https://github.com/JohnZolton) | [zolton.xyz](https://zolton.xyz)

---

## EDUCATION

The University of Akron School of Law - *Juris Doctor* 2023 (Honors Scholar, Dean's List)

The University of Akron - *B.S. Biomedical Engineering* 2020 (Honors, Varsity Track & Field)

## EXPERIENCE

### Machine Learning Engineer – Trunk Tools

Sept 2025 – Present

- **Launched a new Submittal Register product** by productionizing NLP research into a Prefect-orchestrated AWS pipeline, delivering an app-ready service end to end.
- Processed **~3,000-page documents in ~5–8 minutes** through parallelization and pipeline tuning (~4× faster than early runs).
- Improved end-to-end latency by standardizing LLM responses into **token-efficient structured formats**, allowing parallel paths to be merged quickly and deterministically.
- Reduced failed-run rates for a production AI document-review system from ~20–30% to ~≤1% through stabilization, hardening, and targeted bug fixes.

### Machine Learning Architect - Skyward IT Solutions

Jan 2025 – Sept 2025

- **Led MCP initiative:** spearheaded org-wide adoption of Model Context Protocol, enabling tool-augmented LLM agents across multiple teams.
- **Scaled CMS Chat** (most popular tool) from ~50 to 7,000 daily active users by evolving it from rigid path-following to a flexible agentic system.
- Integrated MCP agents with Slack, Jira, Confluence, Linear, SQL/Vector DBs, and live Python execution.
- Developed infrastructure for **LLM inference benchmarking** (Terraform, multi-cloud) with metrics like TTFT + throughput across 25+ open source models.

### Founder/Builder - Socratic Sensei & Patense.ai

2023-2025

- Built two AI SaaS platforms for legal and patent workflows using Next.js, tRPC, Prisma, and LLM pipelines.
- Designed hallucination-resistant, knowledge-grounded LLM systems anchored to user data.
- Achieved 2× inference performance using vLLM with local + cloud privacy-first deployments.
- Improved semantic search speed 300% via prompt design and custom  $O(n \log n)$  retrieval algorithms.

### Patent Attorney - McBee, Moore & Vanik IP, LLC

2023 – 2025

- Secured patent protection across diverse technology areas and managed global prosecution strategies for multiple clients.

## Projects

- Built Gobble, an MCP-powered podcast platform: downloads, transcribes (Whisper), removes ads, and enables semantic + metadata search over transcripts with vector DBs.
- Implemented core neural network architectures from scratch to validate deep learning fundamentals.