

# NITHIN BALAMURUGAN

☎ 647-915-3979 ✉ nithinbala08@gmail.com 🌐 [linkedin.com/in/nithinbala05](https://www.linkedin.com/in/nithinbala05) [nithinbalamurugan.com](https://nithinbalamurugan.com)

## Experience

---

**CIBC** May 2026 – Aug. 2026

*Software Developer Intern* Toronto, ON

- Resolved **100+ production incidents** via root-cause analysis and patch deployment, maintaining **98%+** uptime and cutting repeat issues **25%** across **4+ internal systems** validated on **Azure & Databricks**.
- Automated stuck-file detection via **Excel VBA + Power Automate**, reducing alert latency **95%+**; built **Linux** diagnostic tooling to accelerate root-cause triage.

**Varonova Tech Inc.** Jan. 2026 – Apr. 2026

*AI Software Developer* London, ON

- Refactored procedural video pipeline into **OOP Python (NumPy, OpenCV, FFmpeg)**, deploying **Depth Anything V2** and **DPT-MiDaS** on **Modal** serverless GPUs for scalable depth estimation.
- Built a **3D point cloud** pipeline via **NumPy** backprojection and **4x4 camera pose transforms**, producing consistent scene geometry from monocular depth.

**CognitomeAI** Sept. 2025 – Dec. 2025

*AI Research Platform Intern* London, ON

- Built a **retrieval-augmented chatbot** grounded exclusively in scraped **PhD and peer-reviewed papers**, eliminating hallucinations by restricting generation to verified academic sources.
- Engineered a **FastAPI + GPT-3.5** query pipeline that maps free-form prompts to relevant literature, surfacing citation-backed answers across a live-indexed paper corpus.
- Deployed full stack on **AWS + Docker** with **Next.js** frontend; reduced literature review time from hours to minutes.

**MomentumMind** May 2025 – Aug. 2025

*Data Engineer Intern* Toronto, ON

- Rewrote legacy SQL with **CTEs** and **window functions**, cutting query latency **40%** and enabling real-time **Tableau/Power BI** dashboards.
- Migrated manual Excel workflows to **AWS Pandas** pipelines, saving **5+ hours/week** and improving KPI accuracy.
- Standardized data outputs across **SQL, PDF, and Excel** sources into a unified **ML-ready** schema, enabling downstream model training without manual preprocessing.

## Research / Projects

---

**Tiny Recursive Models on ARC-AGI-1** | [arXiv:2512.11847](https://arxiv.org/abs/2512.11847) | [Research Blog](#)

- Reproduced and analyzed the ARC Prize verification checkpoint across **400+ tasks**, isolating the contribution of ensemble, identity conditioning, and recursion depth.
- Quantified **test-time compute scaling** via 1,000-sample augmentation + majority voting, delivering **+10.75 pp Pass@1** over single-pass inference.
- Proved **puzzle-ID embeddings** are a strict model dependency (accuracy collapses to **0%** without them); showed **94%** of final accuracy achieved at recursion step 1.

**Quick-QA** | *TRM Inference for Enterprise Cost Reduction* | [Research Blog \(Blog 4\)](#)

- Demonstrated **TRMs** as a drop-in LLM alternative for corporate pipelines, leveraging recursive architecture for faster, more accurate inference at lower token cost; solution for reduced corporate expenses.
- Benchmarked vs. **FAISS ANN + Hugging Face** baselines; hierarchical retrieval cut token usage and improved answer accuracy over flat **RAG**.

## Technical Skills

---

**Languages:** Python, C/C++, Java, TypeScript, JavaScript, SQL (PostgreSQL)

**AI / ML:** PyTorch, Hugging Face Transformers, XGBoost, SHAP, ONNX, OpenAI API, Depth Anything V2, DPT-MiDaS

**Data & Cloud:** NumPy, Pandas, AWS, Docker, Modal, Firebase Firestore, PostgreSQL

**Backend & Tools:** FastAPI, Flask, Node.js, Next.js, React, Git / GitHub CI/CD, Jira

**CV / Media:** OpenCV, FFmpeg, OCR (Tesseract, RapidOCR)

## Education

---

**University of Western Ontario** Expected Apr. 2027

*Bachelor of Computer Science* London, ON

- Relevant Coursework: Linux, ARM Assembly, Data Structures & Algorithms, Systems Programming, OOP, Database Systems, Networking