

# Amirreza Sokhankhosh

☎ 431-293-6515   ✉ [amirreza.skhn@gmail.com](mailto:amirreza.skhn@gmail.com)   [in LinkedIn](#)   [GitHub](#)   [Portfolio](#)

## HIGHLIGHT OF RELEVANT SKILLS

---

- **Generative AI & LLMs:** Designed and deployed applications powered by LLMs using **LangChain**, **OpenAI API**, **vector databases (FAISS)**, and **Ollama**, with expertise in **retrieval-augmented generation (RAG)**, fine-tuning (LoRA), and advanced prompting strategies (CoT, ReAct, Zero/Few-Shot).
- **Machine Learning Engineering:** Skilled in developing, training, and scaling ML models with **PyTorch**, **TensorFlow**, **Hugging Face Transformers**, applying best practices in distributed AI, model evaluation, and **MLOps** (Docker, Kubernetes, CI/CD) for production-grade pipelines.
- **NLP & Computer Vision:** Applied deep learning to **text and image data** through projects in **natural language processing**, **OCR**, and **image detection/segmentation**, demonstrating the ability to handle multimodal ML challenges.
- **Full-Stack Development:** Proven track record in building and integrating **end-to-end applications**, combining **React (TypeScript)** for intuitive frontends with scalable backends in **Flask**, **Django**, **NestJS**, and databases such as **PostgreSQL**.
- **APIs & Systems Design:** Strong background in architecting and scaling **inference APIs** and cloud-native solutions, bridging robust backend services with user-friendly interfaces.
- **Cloud & DevOps:** Hands-on with **AWS**, **Docker**, **Kubernetes**, **GitHub Actions**, and CI/CD pipelines to ensure reliable and scalable deployments.
- **Collaboration & Agile Practices:** Experienced in **Agile/Scrum** workflows, cross-functional collaboration, code reviews, and delivering high-quality, production-ready systems in both research and industry settings.

## EXPERIENCE

---

### University of Manitoba

Sep 2023 – Jul 2025

*Graduate Research Assistant*

*Winnipeg, Canada*

- Designed/developed four novel distributed AI architectures, emphasizing secure, scalable & efficient **system design and modular API development** (TensorFlow, PyTorch, Hyperledger Fabric, Express.js/Flask).
- Achieved significant system optimizations: reduced communication overhead by **85.2%** & increased fault tolerance by **62.7%**, demonstrating expertise in building reliable AI systems.
- Authored/contributed to multiple research publications in top-tier **IEEE** venues.

### Bobo App Ltd.

May 2024 – Aug 2024

*Full Stack Developer Intern*

*Winnipeg, Canada*

- Collaborated with a team at **MILA** in developing an efficient **chatbot** responsible for **customer service**.
- Accelerated product development by designing and implementing **RESTful APIs** using Supabase and **PostgreSQL**, enhancing backend functionality and data management.
- Automated data integration by developing a Python script to convert CSV data into executable SQL, significantly streamlining **database population**.
- Ensured seamless **full-stack integration** through close collaboration with the front-end team to align API specifications with user interface requirements.
- Contributed to **agile** workflow efficiency by utilizing the Atlassian suite (**Jira**, **Confluence**) for task management, documentation, and team coordination.

### Nadin Soft (Sadr Group Company)

Jul 2020 – Dec 2020

*Full-stack Developer Intern*

- Drove the successful completion of a critical internal project, addressing key deficiencies in **API security**, **authentication**, and **data modeling** for production deployment.
- Developed robust RESTful APIs with AdonisJS and **Express.js**, implementing **JWT authentication** to secure user access and enable core application functionalities.

- Designed and implemented highly accurate MySQL database schemas for precise tracking and management of employee working and non-working hours, foundational for the project's business logic.
- **Deployed** fully functional features to the **live production server** and created thorough Swagger API documentation, facilitating seamless internal integration and knowledge transfer.
- Contributed to an Agile Scrum development team by actively tracking and managing personal tasks within ClickUp, ensuring efficient progress and clear communication within sprints and **CI/CD pipelines**.

## PROJECTS

### MarkMate | *React, Django, PostgreSQL, OpenAI API, LangChain* | [Link](#)

- Built a full-stack web app for automated assignment grading using **React, Django, and PostgreSQL**.
- Integrated OCR and RAG-based **LLM agents** with custom test-case pipelines to generate consistent feedback.
- Reduced grading time and instructor effort by **≈80%**, improving reliability and scalability.

### Paper Summarizer | *Hugging Face Transformers, Ollama, OCR, LLMs* | [Link](#)

- Engineered an LLM-powered summarization tool for academic papers with **OCR + NLP pipelines**.
- Applied **LLaVA and LLaMA models** for segmentation, entity extraction, and multimodal summarization.
- Generated holistic one-page summaries, improving research efficiency for academic users.

### Unified Image Evaluation Metric | *PyTorch, NumPy* | [Link](#)

- Designed a novel evaluation metric for generative AI, integrating **precision, recall, quality, and diversity**.
- Quantified generalization with **KNN-based scoring** and unified results via harmonic sum.
- Applied as a regularization term to **WGAN-GP and Diffusion models**, improving generalization.

### Digital Twin – FL | *TensorFlow, PyTorch Forecasting, Flask, Express.js* | [Link](#)

- Developed a smart-building **Digital Twin** predicting CO<sub>2</sub>, humidity, and temperature across 76 IoT rooms.
- Implemented **Temporal Fusion Transformers (TFT)** for multivariate time series forecasting.
- Integrated into a **federated learning framework**, ensuring scalability and privacy-preserving analytics.

### Facial Landmark & Boundary Detection | *OpenCV, TensorFlow, Faster R-CNN* | [Link](#)

- Developed a computer vision pipeline for **multi-face detection and landmark localization**.
- Trained dual **Faster R-CNN models** to detect boundaries and extract 68 facial landmarks.
- Delivered robust performance for real-time facial feature extraction.

## EDUCATION

### University of Manitoba

Sep 2023 – Aug 2025

*Master of Science in Computer Science (GPA: 4.4 / 4.5)*

*Winnipeg, Canada*

- **Relevant Coursework:** Security & Privacy, Deep Generative Modeling, Blockchain & Distributed Systems: A+

### K.N. Toosi University of Technology

Sep 2018 – Feb 2023

*Bachelor of Science in Computer Engineering*

## TECHNICAL SKILLS

**AI / Machine Learning:** TensorFlow, PyTorch, Flower, Numpy, Pandas, OpenCV, Transformers, LangChain, LangGraph, Scikit-learn, Keras, Matplotlib, TensorBoard, CUDA.

**Languages:** Python, C++, Java, JavaScript, TypeScript, Node JS, Solidity.

**Cloud & DevOps:** Docker, Kubernetes, Hyperledger Fabric, Git, GitHub, Linux, CI/CD, AWS.

**Databases:** PostgreSQL, MongoDB, MySQL, Redis, Db2, Supabase.

**Web Frameworks:** *Back-end:* NestJS, Django, ExpressJS, Flask, AdonisJS. *Front-end:* React, Tailwind

**Tools & Methodologies:** Jira, Confluence, Swagger, ClickUp, Agile, Scrum.

## CERTIFICATES

### Inferring Causal Effects from Observational Data

*University of Pennsylvania* | [Credential](#)

### Reinforcement Learning Specialization

*University of Alberta* | [Credential](#)