Research Assistant (AI/ML Developer) at Carleton Immersive Media Studio

We are currently seeking people to fill positions (part-time or full-time) as developers in the CIMS lab. If you are interested in a position at CIMS, please submit your CV and portfolio/link by email to the following:

To: Professor Stephen Fai, Director, CIMS, sfai@cims.carleton.ca

With CC: Laurie Smith, Research Operations, CIMS, lsmith@cims.carleton.ca,

Lara Chow, Associate Director, CIMS, lchow@cims.carleton.ca,

Nicolas Arellano, Team lead, CIMS, narellano@cims.carleton.ca

Background & Description of Organization:

Carleton Immersive Media Studio (CIMS) is a Carleton University Research Centre (CURC) affiliated with the Azrieli School of Architecture and Urbanism and the Department of Civil and Environmental Engineering. Our research addresses the development of hybrid workflows that both acknowledge the invisible measures of architecture and animate the visible world of construction. Our work uses established and emerging digital technologies to explore and support architectural rehabilitation, heritage conservation and built-asset management. We are advocates for the thoughtful and critical integration of computing and information technologies within existing cultural and disciplinary contexts.

Over the past decade, CIMS has developed an international reputation — working with public, private, and not-for-profit partners — to create a culturally rich and diverse portfolio of projects. We are engaged in five streams of research: Digitization, Modelling, Digitally Assisted Fabrication, Digitally Assisted Storytelling and Digital Twins. Information on past and current CIMS projects is available online at https://www.cims.carleton.ca.

Job Description: Research Assistant (AI/ML Developer)

We are seeking a skilled AI/ML Developer to join our multidisciplinary team working on an innovative open-source Digital Twin platform for deep energy retrofits across Atlantic Canada. The ideal candidate will be responsible for designing and implementing an agentic AI system that enables non-technical users to interact with complex building stock datasets through plain-language queries, removing the need for SQL expertise or specialized software.

Working with a multidisciplinary team in one or more of the five research streams identified above, you will carry out the following responsibilities:

- Design, develop, and maintain agentic AI systems using Python and modern LLM orchestration frameworks (e.g., LangGraph)
- Assist in the design of the core agent architecture, including orchestration logic and specialized agents for query interpretation, task planning, and response synthesis
- Develop prompt engineering and tool-use frameworks for large language models
- Develop Text-to-SQL pipelines for translating natural language queries into structured PostgreSQL/PostGIS queries
- Implement Retrieval-Augmented Generation (RAG) architectures for retrieval over technical documents, retrofit guidelines, and policy materials
- Implement accuracy guardrails, hallucination mitigation strategies, and response validation mechanisms

- Develop evaluation frameworks covering technical performance, reasoning reliability, usability, and robustness
- Conduct user testing of the AI Agent with partner stakeholders and iterate on findings
- Create and maintain APIs for seamless communication between the agent and frontend interfaces
- Collaborate with frontend, backend, and data engineering team members to integrate the agent into the Digital Twin platform
- Develop reusable code and libraries for AI/ML workflows
- Research solutions to technical problems related to LLM reasoning, agent design, and natural language interfaces
- Test and improve existing AI systems through observability tooling, logging, and tracing
- Maintain thorough documentation of agent design, evaluation results, and future development recommendations

Your core skills must include:

- Proficiency in Python and familiarity with the modern AI/ML ecosystem
- Experience working with large language models (LLMs) through APIs (OpenAI, Anthropic, open-weight models)
- Familiarity with prompt engineering, tool calling, and agentic workflow patterns
- Familiarity with LLM orchestration frameworks (e.g., LangGraph, LangChain, CrewAI, LlamaIndex)
- Understanding of Retrieval-Augmented Generation (RAG) concepts, including embeddings and vector databases
- Familiarity with SQL and relational databases (PostgreSQL preferred)
- Ability to design and implement RESTful APIs
- Experience collaborating using GitHub for version control and project management
- Familiarity with open-source libraries and their management
- Understanding of structured and unstructured data

It would be great if you had the following skills:

- Experience building Text-to-SQL systems or natural language interfaces over structured databases
- Familiarity with vector databases (pgvector, Chroma, Weaviate, Qdrant)
- Experience with LLM evaluation frameworks (RAGAS, DeepEval) and benchmark design
- Experience with observability and tracing tools for LLM applications (LangSmith, Langfuse, Arize)
- Familiarity with PostGIS or geospatial data and queries
- Familiarity with web development frameworks such as Next.js or FastAPI
- Comfortable with Linux
- Strong knowledge of Docker and managing containerized environments
- Knowledge of cloud platforms and services (Fullhost, Azure, AWS, GCP)
- Understanding of fine-tuning, model evaluation, and ML experiment tracking
- Familiarity with frontend technologies (React.js) for better collaboration

- Experience with version control systems like Git

We are looking for a candidate who:

- Is independent and has a willingness to learn new AI technologies and concepts
- Possesses strong problem-solving and analytical skills
- Has an interest in research-oriented AI development and applied LLM systems
- Can communicate complex AI/ML concepts effectively with team members and stakeholders
- Has a growth mindset and is receptive to changes in the rapidly evolving AI technology landscape
- Demonstrates attention to detail and a commitment to writing clean, scalable code
- Approaches AI development with rigour, including careful attention to evaluation, reliability, and failure modes
- Can work efficiently both independently and as part of a team on AI-related projects
- Is eager to learn new skills and technologies in the field of AI/ML and agent development
- Has excellent communication skills and is able to explain technical AI details effectively
- Conduct troubleshooting, debugging, and performance optimization of AI systems
- Is able to work independently and in a team environment
- Is eligible for federal security clearance in Canada
- Has demonstrated experience working in multi-disciplinary teams on AI- or data-centric projects
- **Must be currently enrolled as an undergraduate or graduate student, or a recent graduate (within 1 year of graduation)**

Please include a link to your GitHub page with your application and links to any deployed projects/websites you have developed/contributed to.

Join our innovative team and contribute to the development of a cutting-edge open-source Digital Twin platform. If you're passionate about applied AI, large language models, agentic systems, and open-source projects, we'd love to hear from you!