

# Ehsan Hamzei

## Data Scientist/Software Developer

Hamzei.ehsan@gmail.com | +61-450087660 | Melbourne, VIC

<https://www.linkedin.com/in/ehamzei/> | <https://github.com/hamzeiehsan>

## SUMMARY

I am a **Data Scientist and an Experienced Software Developer** with a **Ph.D. in Information Systems**. I have professional team-work experience leading and working with data specialists, software developers and researchers. I bring value to projects as a trusted data scientist who can take responsibility and ownership of projects from the initial pitch to the final delivery. I enjoy teamwork and have the experience of working on various machine learning projects including **natural language processing, time series analysis and spatial computation**. I have been a key member of several successfully delivered ARC funded projects in academia, and data science and business analytics projects in industry.

## KEY SKILLS

- Deep knowledge of **machine learning** (regression, classification, clustering, transfer learning), **deep learning** (RNN, LSTM, transformers), **statistics** (probability theory), and **data analysis** with experience in **text, image, spatial data, and time series**.
- Professional programmer in **Python, R, Java**, with coding experience in Linux environments.
- Extensive experience with common data science libraries (NLTK, Scikit-learn, SpaCy, SciPy, NumPy, Pandas, dplyr, Tidyverse, caret) and machine learning frameworks (TensorFlow and PyTorch).
- Experience in mining large and complex datasets using **SQL** (Oracle, Postgres), **Cypher** (Neo4J), **SPARQL** (RDF storage) and **pySpark**.
- Professional experience in using and extending **open-source technologies** (Apache Kafka, Hadoop, Zookeeper, Lucene and Solr).
- Familiar with cloud computing (Amazon S3, Sagemaker), data warehousing (Oracle Data Warehouse, DataBricks), model development and deployment technologies (MLflow).

### Soft skills:

- Versatile and flexible data scientist capable of thriving independently, collaborating within a team, or taking up leadership roles.
- Creative and analytical thinker, approaching problems with a novel perspective.
- Proficient in verbal and written communication and adept at cultivating strong interpersonal connections.

## RELEVANT EXPERIENCE

### Senior Data Scientist

January 2024 - Present

nCino Inc. Melbourne, AU

#### Key Responsibilities

- Working in cutting-edge data science projects including vision and language model training and development.
- Automating the process of information extraction from document
- Facilitating the communication between engineering and data science team

#### List of Projects

- **Automated Information Extraction From Documents**

**Techniques:** Vision Transformers, LLM

**Technologies:** Databricks, Python, HuggingFace Transformers, Pytorch, Lightning, PySpark

### Research Fellow - Data Scientist

July 2021 – January 2024

University of Melbourne, Parkville, VIC

#### Key Responsibilities

- Leading and working in several data science projects.

- Developing and using local large language models in analysing a repository of cultural heritage documents from **Heritage Victoria**.
- Leading a team of 4 researchers at the University of Melbourne to develop an automatic ETL pipeline to create spatial pedestrian network from public datasets for an **ARC** project.
- Identifying, analysing, and interpreting patterns in massive time series datasets from **nbn** Australia datasets.

### List of Projects

- **Using Large Language Models (LLMs) in Analysing Cultural Heritage Reports:**
  - **Leading** a team of 6 researchers from University of Melbourne and RMIT to design an information extraction and generation workflow and a **chatbot** for facilitating the manual process of analysing heritage reports.
  - Using local open-source large language models and information retrieval and geocoding techniques in extraction process.
  - Evaluating the extracted information and generated responses (**82%** relevant responses, **0.59 km** error in finding location of the heritage sites)  
**Techniques:** Information retrieval, word embeddings, LLM text generation, Prompt engineering.  
**Technologies:** Python (LangChain, Flair, Flask, Spacy), ChromaDB (vector database) and SQLite (relational database), Linux.

## **Data Engineer - Research Assistant**

**November 2019 – June 2021**

University of Melbourne, Parkville, Australia

### Key Responsibilities

- Work as a data engineer to design an automatic workflow that combines several environmental, social and health datasets in Australia.
- Maintain the system and present a demo to researchers at University of Melbourne.

### Project

- **Creating a coherent spatiotemporal database using available datasets in Australia (MSSI Funded Project):**
  - Collect and clean heterogeneous datasets from the state data portals and local government websites including sensor datasets for environmental variables, market datasets such as energy datasets from OpenNEM, health dataset from Health Victoria, social datasets such as crime dataset from Crime Statistics Agency.
  - Design and implement an ETL workflow to automatically extract, transform and load the data into the spatiotemporal dataset.  
**Techniques:** ETL, tabular data processing, time series processing.  
**Technologies:** Python (sqlAlchemy, Pandas, Numpy, Flask, SciPy), Postgres and PostGIS (relational database).

## **PhD Researcher**

**July 2017 – May 2021**

University of Melbourne, Parkville, Australia

### Key Responsibilities

- Developed machine learning models for automatically answering spatial questions.
- Communicated and collaborated with researchers in Australia, USA, Belgium and Netherlands and published several peer-reviewed publications.

### Achievements

- Developed a question answering model using **deep neural networks** (AllenNLP, transformers) and transfer learning using **PyTorch** in Python.
- Designed an automatic workflow to translate spatial natural language questions to GeoSPARQL and Spatial SQL queries using pre-trained DNN models and NLP techniques (dependency and constituency parsing). Implemented in Python using PyTorch, AllenNLP, SciPy, AnyTree and PySolr. Improving the accuracy of SOTA models to **79.5%**.
- Performed analysis on a massive corpus of questions from Microsoft (MS MARCO, with more than **1M questions**) and extracted insightful patterns about the questions and relevance of human-generated answers using R and Python.
- Achieved strong communication skills by writing a PhD thesis and research papers and presenting in international conferences and universities.

## **Data Engineer and Programmer**

**April 2014 – June 2017**

Mobin Information Technology Research Centre, Tehran, Iran

## Key Responsibilities

- **Developing** software and APIs using **Java Core and Java EE** technologies in Enterprise Service Bus (ESB) project.
- Guiding a team of three SQL developers in a Business Analytics Project while facilitating seamless communication between the data engineering and data science teams.

## Achievements

- Developing a ESB solution using Java and open-source technologies Apache ActiveMQ, Camel, Zookeeper.
- Designing workshops and tutorials for clients to define their integration scenarios using the developed ESB platform.
- Implementing **RESTful webservice**s for data integration using JAX-RS.
- Dependency management and project deployment using Apache Maven and Karaf.
- Using Object-Relation-Mapping (ORM) in writing database services (JPA, Hibernate and Spring Data).
- Developing and maintain a data engineering workflow from Massive Banking Online Transactional Processing Database (OLTP) and load coherent data into an Online Analytical Processing Database (OLAP) using SQL, Oracle PL/SQL, and Oracle Data Integrator (ODI).
- Presenting the pipeline procedure to different clients and demonstrating its capabilities.

## Side Projects

July 2015 - present

- **Hackathon Project:** Leading a small team for a hackathon at Spatial Data Science Symposium 2021 on geocoding textual location descriptions (**winner:** <https://geocollaboratory.massey.ac.nz/author/kmstock/page/3/>).
- Designing **workshops and tutorials** for international conferences – Conference of Spatial Information Theory 2022 and GIScience23 on GeoAI and GeoQA.
- **Master project:** Designed and developed a prototype spatial search engine using Java and Apache Solr, Oracle Spatial, GeoTools and StanfordNLP to support complex spatial queries in keyword-based search.
- Lead group of tutors and supervise master and PhD students at University of Melbourne (2021 – present).
- Student representative of human Ethic Committee at University of Melbourne (2018-2020)

## EDUCATION

### The University of Melbourne, Australia

July 2017 – May 2021

Doctor of Philosophy (PhD), Information Systems (Spatial)

**Awards:** Melbourne Research Scholarship for 3.5 years.

**Thesis Topic:** Place-related question answering: From questions to relevant answers.

### University of Tehran

September 2013 - September 2016

Master of Science (M.Sc.), Spatial Information Systems

**GPA:** 18.29/20 (academic awardee)

**Thesis Topic:** Spatial search engines: Automating spatial query processing.

### K.N. Toosi University

September 2009 – September 2013

Bachelor of Science (B.Sc.), Geomatics Engineering

**GPA:** 17.85/20 (academic awardee)

**Thesis Topic:** Design and implementation a fire detection algorithm using MODIS satellite imageries.

## HOBBIES

Hiking, Badminton, Gaming, Cooking, Reading.

## Reference

Available upon request.