

AMIN FARVARDIN

A Team Player Who Embracing Challenges to Achieve Project Success

Fields of interest: Architect Data / Data Scientist/Engineer

Current position: R&D Team Leader at HELKINZ (Permanent Position)

✉ m.farvardin@hotmail.com

☎ +33 (0)7 69 11 12 20

📍 Courbevoie, France

in linkedin.com/in/FarvardinA

🔗 lamsade.dauphine.fr/~afarvardin

● leetcode.com/u/farvardin

🐙 GitHub.com/afarvardin



WORK EXPERIENCE AT HELKINZ

R&D Team Leader - Data Scientists

Project/Product Manager

📅 Dec 2021 - Present

📍 Paris France

Contract Generation Using Generative AI

Developed a cutting-edge Generative AI system leveraging:

- Retrieval-Augmented Generation (RAG) techniques
- Vector and knowledge base databases
- Clustering algorithms (DBScan, HDBScan)
- Semantic analysis methods (LDA, LSI)

This system trains a document-based ML model capable of:

- Analyzing a repository of 1000+ legal documents
- Selecting the optimal contract template
- Auto-populating the template with relevant company and personal information

Result: Streamlined contract creation process, significantly reducing time (from days to minutes) and potential errors in legal document preparation.

Tasks:

- Engineered and deployed an ML model utilizing DBScan, HDBScan, LDA, and LSI for semantic search and document clustering
- Leveraged state-of-the-art Large Language Models (LLMs) including LLama-2/3, Mistral AI, and GPT-3/4 via AWS Bedrock for intelligent contract generation
- Implemented sophisticated prompt engineering techniques to optimize LLM performance and accuracy in contract creation
- Developed a novel ML-based approach for suggesting, editing, and rearranging contract sections based on historical data analysis
- Cost Optimization of a scalable AWS architecture to support real-time AI/ML operations, by deploying the custom models (LLMs & ML model) on EC2 instances
- Integrated multiple data sources with AI/ML pipeline, incorporating AWS Knowledge Base, external APIs, and OCR technology for comprehensive data gathering
- Successfully delivered an end-to-end AI-powered solution, from Proof Of the Concept to MVP; aims to revolutionizing the legal contract generation process

Sales Predictions Project

Developed an innovative SaaS solution for small to medium-sized businesses in the food service industry to reduce waste by accurately predicting daily food requirements. By considering:

- Sales forecasting based on historical data
- Recipe ingredient planning and optimization
- Conducted comprehensive research on technical specifications and business requirements
- Developed proof of concepts for the MVP phase
- Designed a cost-effective, scalable architecture on AWS
- Led data science and engineering teams using Scrum methodology
- Developed proof of concepts for the MVP phase
- Designed a cost-effective scalable architecture on AWS, and led the data science and engineering teams using the Scrum methodology to implement the project.

The result is a comprehensive system that enhances operational efficiency, improves customer satisfaction, and promotes sustainable business practices in the food service industry.

Real-Time Banking transaction Fraud detection

Objective: Developed and implemented a real-time fraud detection and prevention system using Big Data, Graph, and Neural Network architectures on AWS Cloud.

Key Responsibilities:

- Design and architect the solution
- Led data science and engineering team as product/project manager using Scrum methodology
- Implemented CI/CD practices for data model development

Technical Highlights:

- Data Processing: Transformed CSV to Parquet, utilized AWS Crawler and Glue for graph structure conversion
- Model Development: Built GNN model using PyTorch and Deep Graph Library (DGL)
- Cloud Infrastructure: Leveraged AWS services including Lambda, SageMaker, Neptune, ECR, ECS, Fargate, SQS, DocumentDB
- Real-time Visualization: Designed architecture using React.js, CloudFront, and AppSync with GraphQL

Key Achievements:

- Implemented real-time fraud detection with high confidence based on behavioral analysis
- Reduced model size by one-third through repackaging with AWS ECS and Fargate
- Created a scalable in-service endpoint using AWS SageMaker
- Designed end-to-end pipeline from data ingestion to visualization of fraud prediction results

Technologies Used: AWS (Lambda, Crawler, Glue, Athena, SageMaker, Neptune, ECR, ECS, SQS, DocumentDB, DynamoDB, Kinesis, AppSync, API Gateway, S3, CloudFront, Step Functions), PyTorch, PySpark, React.js, GraphQL, GitHub, VS Code, AWS CLI

Methodology: Scrum

Skill Sets Gained in the Current Position

- **Agility:** Well experienced in implementing solutions iteratively in an agile environment (Scrum)
- **CI/CD:** Azure DevOps Pipeline, AWS CodeCommit, AWS CodePipeline, GitHub Actions
- **ETL process:** Spark & PySpark data processing (in batch & stream), Delta Lake, AWS Crawlers, AWS Data Catalog, & AWS Glue, AWS Redshift (Data warehouse)
- **Containerization apps:** Dockers/DockerHub, Amazon Elastic Container Registry (AWS ECR) & Amazon Elastic Container Service (AWS ECS)
- **Automation / Scheduling / Monitoring:** Terraform, Azure DevOps Pipeline, Linux Shell Script, AWS Step Functions, AWS CLI, Powershell, WSL, Apache Airflow
- **Data modeling:** MS SQL, MySQL, NoSQL databases, HDFS, MongoDB, AWS DynamoDb, AWS DocumentDB, AWS RDS, AWS Neptune (highly connected datasets), AWS S3
- **Big data:** Streaming/Processing (in Batch & Stream & near Real-time), Batch / Micro-Batch Processing, Apache Kafka
- **Documentation:** technical and business level
- Design and implement a near to a real-time streaming pipeline
- MLOps on AWS SageMaker
- ML & GNN modeling
- Data Scrapping

R&D responsibilities:

- Define projects by analyzing the market demand
- Architecting and designing the relative product on the cloud (AWS) to answer the market demand based on the state of the art
- Proposing technologies and methods for implementation
- Optimizing the projects costs and performance
- Keep them up-to-date relying on the state-of-the-art
- Supervising apprenticeships to learn and contribute to real projects and guide them to write their thesis and defend their masters

WORK EXPERIENCE AT LAMSADE LAB (UNIVERSITÉ-PARIS DAUPHINE)

Data Scientist/Data Engineer in Data Science Pole

Data Streaming, Data Processing, Architecture Big Data

Nov 2016 - Dec 2020

Paris, France

Scalable Process of Streaming Massive RDF Data

- Designed, developed, and implemented a novel indexing technique for RDF saturation in Spark Streaming.
- Implemented lightweight in-stream metadata for data processing.
- The proposed method identified the relationship between the previously received data and the current data.
- Incrementally Saturate Streaming RDF
- Memory usage reduced by approximately 40
- Optimizing the Spark Processes and Partition
- Achieve an approximately 17x reduction in execution time.
- Responsible for maintaining and setting up a 10 nodes on-premise cluster

Spark & Spark Streaming

Kafka

Scala

Java

Python (Pandas, NumPy)

Hadoop Yarn

HDFS

Azure Databricks

AWS EMR/S3

Apache Airflow

WORK EXPERIENCE AT GERIICO LAB - UNIVERSITÉ DE LILLE III

Full Stack Engineer at Text Mining Team

Text mining - NLP

Oct 2015 - Oct 2016

Lille, France

Identification and analysis of land studies in TERRE-ISTEX project

- Identification and analysis of land studies
- Annotate spatial text geographically
- Querying annotated spatial text

Text mining (GATE & GATE Embedded)

NLP

RESTful API

Java

Spring MVC

Elasticsearch

SQL

NoSQL

JS, HTML, CSS

INTERNSHIPS & VOLUNTEER EXPERIENCE

Full Stack Engineer & Data Visualizan

Text mining - NLP - Azure Cloud

Feb 2012 - Oct 2016

France & India

Full Stack Engineer

Text Mining (Internship)

April - Sep 2015

Montpellier, France

Semantic annotation associated with land development projects

Text mining (GATE & GATE Embedded)

NLP

RESTful API

Java

Spring MVC

SQL

JS, HTML, CSS

Data Visualization

SIG L-R (Volunteer)

Feb - April 2015

Montpellier, France

Utilizing the registered participants' data to show what we can do with even none relevant data and how much knowledge is potentially hidden in data

D3.js

JSON

JS

Full Stack Engineer

Text Mining (Master Internship)

📅 Jan 2014 – Jan 2015

📍 Montpellier, France

Semantic web services for opinion mining from land-use planning documents

- Annotate text by creating NLP rules
- Prepare a standalone pipeline
- Transform the standalone pipeline into a website application by utilizing GATE Embedded
- Implement a web tool for analyzing land-use planning documents

Text mining (GATE & GATE Embedded)

NLP

RESTful API

Java

ASP.NET MVC - C#

Spring MVC

SQL

JS, HTML, CSS

Web Development on Azure Cloud

Building a RESTful web service (Internship)

📅 Feb 2012 – July 2012

📍 Hyderabad, India

Design and implement a phone book as a web service on Azure

ASP.NET

SQL

SOAP

Azure

WORK EXPERIENCE AT PETROLEUM INDUSTRY

Software Engineer & Back-end developer

Desktop and Web Applications

📅 June 2004 - July 2008

📍 Shiraz Iran

Automation of office systems

Software Engineer

Petroleum Industry (CDD)

📅 Sep 2006 – July 2008

📍 Shiraz, Iran

- Designed, implemented, and maintained mission-critical software systems for logistics and repair operations in Iran's petroleum industry.
- Developed custom office automation solutions to streamline workflow processes and improve operational efficiency.
- Created intuitive user interfaces for logistics management, reducing training time by 30% and increasing user adoption.
- Implemented a robust database system for tracking repair schedules and inventory, resulting in a 25% reduction in equipment downtime.
- Collaborated with cross-functional teams to gather requirements and ensure software solutions met industry-specific needs
- Utilized SQL for database management and optimization, improving query performance by 40%.
- Implemented data analytics and reporting features, providing management with actionable insights for decision-making.
- Ensured software compliance with petroleum industry regulations and security standards
- Conducted regular system maintenance and updates, resulting in 99.9% up time for critical applications

C#.NET

SQL

Excel

VB.NET

Web Site developer

Back-End Web Developer

ESG Co. Ltd. (CDD)

📅 June 2004 – June 2006

📍 Shiraz, Iran

- Utilizing DotNetNuke (DNN) as a Content Management System (CMS) Platform
- Developing individual modules (weather, calendar, etc.)
- Packaging the module to make them installable on the raw DNN
- Resolving the library conflicts during packaging and installation
- Maintaining the website after selling
- Working in a team of 5 consisting of 2 web designers and 2 web developers including me

ASP classic

ASP.NET

MS SQL

HTML, CSS, JS

CMS (dnn)

EXPERTISE* AND INTERESTS**

Data Architecture*

- Batch Processing, Stream Processing, Micro-Batch Processing
- Lambda and Kappa Architecture
- Knowledge management
- Semantic web – ML algorithms

Developments**

- Text mining (NLP)
- Software Engineering
- Cloud Architecture (AWS)

EDUCATIONS

Ph.D. in Streaming Big Data (Université-Paris Dauphine)

Title: Scalable Saturation of Streaming RDF Data

📅 Nov'16 – Jan, 2021

📍 Paris, France

Pioneered an innovative solution for real-time reasoning over large-scale RDF graph data streams:

- Developed the first-of-its-kind incremental saturation algorithm for RDF data, significantly outperforming existing bulk saturation methods
- Engineered a sophisticated system to identify relevant data fragments for processing, optimizing computational efficiency
- Implemented the solution using big data platforms, demonstrating scalability for handling massive RDF datasets
- Incorporated RDF schema and OWL-Horst constraints to infer implicit RDF triples, enhancing data completeness and query accuracy
- Achieved substantial performance improvements, reducing processing time by 17x compared to traditional approaches
- Memory usage reduced by approximately 40
- Published findings in DBPL 2019: "Streaming saturation for large RDF graphs with dynamic schema information"
- Potential applications include real-time semantic web analysis, dynamic knowledge graphs, and intelligent data integration systems

Master (Information Technology, JNTU)

Project: Semantic web services for opinion mining from land-use planning documents

📅 Sep 2012 – Oct 2014

📍 Hyderabad, India

Grade: First Class with Distinction

Bachelor (Computer Engineering, JNTU)

Project: A Phone-Book as a web service on Azure

📅 Sep 2008 – Jun 2012

📍 Hyderabad, India

Associate's Degree

Computer Software Engineering

📅 2001 – 2004

📍 Shiraz, Iran

PUBLICATIONS (TOP 5)

1. Farvardin MA., Colazzo D., Belhajjame K., Carlo Sartiani. "Scalable Saturation of Streaming RDF Triples", TLDKS 2020 (Journal).
2. Farvardin MA., Colazzo D., Belhajjame K., Carlo Sartiani. "Streaming saturation for large RDF graphs with dynamic schema information", DBPL 2019, Phoenix, AZ, USA.
3. Kergosien E., Farvardin MA., et al. "Automatic Identification of Research Fields in Scientific Papers", LREC 2018, Japan.

4. Farvardin MA., Kergosien E., Roche M., Teisseire M. "A Webtool for Analyzing Land-use Planning Documents.", ISWC (Posters & Demos) 2015.
5. Farvardin MA., Kergosien E. "Senterritoire for Spatio-Temporal Representation of Opinions", NLDB 2014.

TEACHING & MENTORING

SIU Adjunct Professor

Schiller International University - Paris Campus

Computer science subjects for Master and Bachelor level

📅 Sep 2022 - Present
📍 Paris, France

Instructor of Big Data Architect

Paris-Dauphine | Tunis-Dauphine University

Continue the collaboration between Paris-Dauphine and Tunis-Dauphine universities; aims to train and award valid certificates to the participants.

- Data Processing (Data Lake / Delta / LakeHouse)
- Stream Processing (Lambda and Kappa Architectures)

📅 2021 - Present
📍 Paris, France

Training and Mentoring Data Scientists & Data Engineers & a Cloud Engineer

l'Université de Paris Cité - l'Université de Paris 8 - & HELKINZ Ltd. Co.

So far (Sept 2024), six Apprenticeship students have been supervised by me on the real projects under my supervision.

The projects were about:

- Contract Generation
- Sales prediction
- Real-time fraud detection

Which all have been designed and implemented on the public clouds.

📅 Dec 2021 - present
📍 Paris, France

Instructor of Big Data Architect

Paris-Dauphine | Tunis-Dauphine University

A collaboration between Paris-Dauphine and Tunis-Dauphine universities; aims to train and award valid certificates to the participants. Participants are from master's students and data scientists to CEO of companies.

- Data Processing (Data Lake / Delta / LakeHouse)
- Stream Processing (Lambda and Kappa Architectures)

📅 June - July 2021
📍 Paris, France

Lecture and Research Assistant

LAMSADE, Université Paris-Dauphine

- Systemes de Langages de Programmation pour le "Big Data"
- Large Scale Machine Learning in Apache Spark
- Mise à niveau en Java
- Database (RDBMS)

📅 Sep 2019 - July 2020
📍 Paris, France

Lecture and Research Assistant

LAMSADE, Université Paris-Dauphine

- Systemes de Langages de Programmation pour le "Big Data"
- Large Scale Machine Learning in Apache Spark
- Mise à niveau en Java
- Database (RDBMS)

📅 Sep 2019 - July 2020
📍 Paris, France

Instructor of C Language course

Shiraz University

Teaching and preparing 45 students for an entrance exam of Master level.

📅 Sep 2003 - July 2004
📍 Shiraz, Iran

STRENGTHS

- Development Languages, Frameworks, & DB

D: Scala JAVA Python (DS Libraries) C# & VB.NET C Language

F: Spark Hadoop Yarn Kafka Elasticsearch AWS Services Airflow Docker ASP.NET MVC
Spring MVC

DB: NoSQL (HDFS, MangoDB, AWS S3 & DynamoDB, etc.) MySQL, MSSQL RDF/XML Apache TinkerPop

- Skills

Stream & Batch Processing (Big Data) Scrum Text mining (NLP) GATE (GUI, Embedded) Linux
Git/GitHub/GitLab Spring MVC ETL concepts Design Patterns RESTful & SOAP d3.js L^AT_EX
JS, jQuery HTML5, CSS3 IIS, Glassfish 4.0

SOFT SKILLS

- Leadership & Teamwork: Team collaboration, empathy, leadership, work ethic
- Personal Attributes: Adaptability, flexibility, responsibility, self-motivation
- Professional Skills: Time management, problem-solving, active listening
- Growth Mindset: Continuous personal development

FUNCTIONAL SKILLS

- Product/Project manager: Leading a team of data scientists and data engineers to deliver products on time and the right market requirement
- As a research engineer, I worked and collaborated with 11 researchers (text mining field) in France (Montpellier, Pau, Lille)
- Team Work: Agile methodology/DevOps, Scrum, GitHub, Slack
- Data Instructor

HONERS & AWARDS

Full PhD Scholarship at Université-Paris Dauphine
Minister of High Education of France

📅 2016
📍 Paris, France

Master II Scholarship in CIRAD
Minister of High Education of France

📅 2014
📍 Montpellier, France

MCSD
Web Application (Credential ID: E326-9247)

📅 Jan 2013
📍 Hyderabad, India

LANGUAGES

English
Persian
French



REFERENCES

Dario COLAZZO

Professor at Université Paris-Dauphine

@ dario.colazzo@lamsade.dauphine.fr

🌐 <https://www.lamsade.dauphine.fr/~colazzo/>

☎ +33 1 44 05 44 09

Khalid BELHAJJAME

Maître de conférences at Université Paris-Dauphine

@ khalid.belhajjame@lamsade.dauphine.fr

🌐 <https://sites.google.com/site/kbelhajjame/>

☎ +33 1 44 05 43 09

Eric KERGOSIEN

MAÎTRE DE CONFÉRENCES at Université Lille-III

@ eric.kergosienuniv-lillefr

🌐 <https://pro.univ-lille.fr/eric-kergosien/>

☎ +33 6 61 80 62 22

Maguelonne TEISSEIRE

Research Professor - INRAE

TETIS Lab. Departement of Information System

@ Maguelonne.TEISSEIRE@inrae.fr

🌐 <https://textmining.biz/Staff/Teisseire/>

☎ +33 4 67 54 87 62