

# Naser Ahmadi

Location: Tehran, Iran | Mobile: +989171015903

Email: [ahmadi.naser68@gmail.com](mailto:ahmadi.naser68@gmail.com)

## DATA SCIENTIST

---

I am a Data Scientist who specializes in natural language processing. I have worked on many projects using state-of-the-art technologies like transformers in various NLP fields, including ChatBot, Question Answering, Summarization, and more. This has given me a solid understanding of different NLP fields. In addition to NLP, I have also worked on projects related to graph analysis, knowledge graphs and other data science fields.

## TECHNICAL SKILLS

---

<b>Data Science</b>	: Machine Learning, Deep Learning, Natural Language Processing, Knowledge Graphs, RDF, Graph Neural Networks, Data Analysis and Visualization
<b>Languages</b>	: Python, Java, SQL, SPARQL, HTML, CSS
<b>MachineLearning</b>	: Numpy, Scikit-learn, Scipy, Pandas, Fonduer
<b>DeepLearning</b>	: Pytorch, Huggingface Transformers, Hugging face's Datasets, Hugging face's Evaluate, Hugging face's Tokenizers, Pytorch Lighting, Gensim
<b>Language Model</b>	: LangChain, OpenAI, PEFT, RAG
<b>Graph Analysis</b>	: OpenLink Virtuoso, PyG, Nebula Graph, Networkx
<b>Plotting</b>	: Matplotlib, Plotly
<b>Web Programing</b>	: Flask, FastAPI
<b>Web Scraping</b>	: Scrapy, Newspaper, Requests, Beautiful Soup, Urllib
<b>Databases</b>	: Redis, ElasticSearch, MySQL, SQLite
<b>Tools</b>	: Git, $\text{\LaTeX}$ , Docker
<b>Skills</b>	: Design Patterns, Test-driven Development, Object Oriented Programming, Restful Services
<b>OS</b>	: Ubuntu, Windows

## SELECTIVE WORK EXPERIENCES

---

<b>Data Scientist</b> <b>Hamrah-e-Avval (MCI)</b>	Jan 2022 – Present Tehran, Iran
<ul style="list-style-type: none"><li>• <b>LLM.</b> Developed a large language model for Persian language [Ongoing]</li><li>• <b>Chatbot.</b> Developed a custom chatbot for MCI FAQs, using LangChain and OpenAI API alongside prompt engineering methods [Ongoing]</li><li>• <b>MTS.</b> Implemented a multi-document summarization model based on mBart and Vespa</li><li>• <b>QG.</b> Developed a question generation model using mT5 on real user data to improve our query suggestion</li><li>• <b>QA.</b> Implemented and Developed Extractive and Generative text-based and table-based question answering models</li><li>• <b>Infobox.</b> Designing and implanting a knowledge graph and a Top Stories Suggestion module</li><li>• <b>PQ.</b> Designing and developing search engine's web graph and page and domain quality algorithms for over 5 billion web pages ]</li></ul>	
<b>Data Scientist, Part-time (As part of the PhD)</b> <b>KPMG</b>	Oct 2018 – Jan 2022 Berlin, Germany
<ul style="list-style-type: none"><li>* Generation and Curation of an Audit Knowledge graph</li><li>* Relation extraction in structured documents</li><li>* Developing methods Text to Structured Text Matching</li><li>* Topic modeling for (German and English) audit texts by employing algorithms such as lda and gsdmm</li></ul>	
<b>Researcher and Developer</b> <b>TGBSCO</b>	Sep 2017 – Mar 2018 Tehran, Iran
<ul style="list-style-type: none"><li>• Researcher, Developer and Algorithm Designer in WiniGames and Momena projects</li></ul>	

- Designing and developing a Recommender system for application suggestion
- Implementing a NLP module to assign tags to sport news

**Researcher and Developer**  
**ARMAN RAYAN SHARIF RESEARCH CENTER**

Nov 2014 – Mar 2016  
 Tehran, Iran

Implemented different graph analysis plugins for Gephi

Implemented text processing algorithms such as main content extractor, text normalizer, tokenizer, keyword extractor by extending Stanford NLP

## EDUCATION

---

**Sorbonne University**

*Ph.D., Data Science*

France

Oct 2018 – Dec 2021

**Thesis Title.** A Framework for the Continuous Creation of a Knowledge Base System

**Supervisor.** Dr. Paolo Papotti

**University of Tehran**

*M.Sc., Software Engineering*

Tehran, Iran

Sep 2013 – Sep 2016

**Thesis Title.** Fraud Detection in electronic transactions using multi-dimensional social network analysis

**Supervisor.** Dr. MahmoudReza Hashemi

**Shiraz University of Tech.**

*B.Sc., Software Engineering*

Shiraz, Iran

Sep 2008 – Sep 2012

**Thesis Title.** Proposing an Open Source Solution for Organizations Based on Evaluation and Comparison of Existing Organizational Open Source Softwares

**Supervisor.** Dr. Raouf Khayami

## AWARDS

---

- Ranked 18th in National Entrance Exam for Universities in Master of Science.
- Ranked as top 1% in National Competition for University Entrance Exam, Shiraz, Iran, 2008.
- External Reviewer in conferences: Bigdata 2020, EDBT 2021.

## PUBLICATION

---

- **[PhD Thesis]** Ahmadi, Naser. A framework for the continuous curation of a knowledge base system. Diss. Sorbonne Université, 2021.
- **[EDBT/ICDT 2022]** Ahmadi, Naser, Hansjorg Sand, and Paolo Papotti. "Building A Knowledge Graph for Audit Information." EDBT/ICDT Workshops. 2022.
- **[ICDE 2022]** Ahmadi, Naser, Hansjorg Sand, and Paolo Papotti. "Unsupervised Matching of Data and Text." 2022 IEEE 38th International Conference on Data Engineering (ICDE). IEEE Computer Society, 2022.
- **[Wiki Workshop 2021]** Ahmadi, Naser, and Paolo Papotti. "Wikidata logical rules and where to find them." Companion Proceedings of the Web Conference 2021. 2021.
- **[EMNLP 2021]** Saeed, Mohammed, Ahmadi, Naser, et al. "RuleBERT: Teaching Soft Rules to Pre-Trained Language Models." Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021.
- **[JDIQ 2020]** Ahmadi, Naser, et al. "Mining expressive rules in knowledge graphs." Journal of Data and Information Quality (JDIQ) 12.2 (2020): 1-27.
- **[JDIQ 2020]** Ahmadi, Naser, et al. "Rulehub: A public corpus of rules for knowledge graphs." Journal of Data and Information Quality (JDIQ) 12.4 (2020): 1-22.
- **[TTO 2019]** Ahmadi, Naser, et al. "Explainable Fact Checking with Probabilistic Answer Set Programming." Conference on Truth and Trust Online. 2019.

## REFERENCES

---

References Available Upon Request