

## Ali Abbas Baygi

Data scientist at Smartblick

**Resident:** Berlin (permanent residence)

**Languages:** English, German, Farsi (mother tongue)

**E-mail:** [alibaygi@gmail.com](mailto:alibaygi@gmail.com)

**Website:** <https://alibaygi.github.io>



I'm a data scientist with extensive expertise in LLMs, data analysis, algorithm design, and the development of robust machine learning software. Holding a Master's degree from TU Berlin, I possess over 5 years of hands-on experience in modeling intricate systems, alongside a proficiency in comprehending, forecasting, and visually representing complex datasets. Coming from a self-taught background, I understand the difficult and winding path towards becoming a data scientist or machine learning engineer.

I am passionate about deep work and able to quickly master hard things. I immigrated from Iran, where I studied mechanical engineering. Ethical Considerations: Awareness of ethical issues and biases in NLP models, along with strategies for addressing them.



### Expertise

#### Statistics

Confidence intervals, significance estimation, distributions, hypothesis testings, A/B testing and experimental design, p-value interpretation, principal component analysis (pca), kernel density estimation (kde), regression analysis (linear, logistic, etc.), time series analysis, bayesian statistics and inference.

#### Supervised / unsupervised ML & Deep Learning

Gaussian process, k-means, random forest, deep learning, feature importance, anomaly detection, dimensionality reduction, debiasing data, NN, and RNN.

#### Natural language processing and Generative AI with LLMs

Text preprocessing, named entity recognition (NER), sentiment analysis, POS tagging, word2vec, bag of words, TFIDF, transformers and attention mechanisms (e.g., self-attention, multi-head attention), LLMs (GPT-3, Bloom, BERT, RoBERTa, T5), fine-tuning and transfer learning with pre-trained models, RAG (retrieval-augmented generation).

#### Digital Signal processing

Signal analysis, filtering and convolution, discrete fourier transforms, spectral analysis convolution, cross-correlation, noise reduction, feature extraction.



### Skills

Python programming, Linux, WebScraping, Web Service API, Git, Gitlab, CI/CD, NumPy, pandas, matplotlib, sklearn, PyTorch, transformers, nltk, fastapi, streamlit, Keras, Tensorflow, LangChain, SQL, mongodb, elasticsearch, chromaDB, myScale, Application containerization e.g Docker.

### Licenses & certifications

- TensorFlow Developer Professional Certificate
- Generative AI with Large Language Models
- Natural Language Processing in TensorFlow
- Convolutional Neural Networks in TensorFlow
- Neural Networks and Deep Learning
- Sequence Models
- Machine Learning Data Lifecycle in Production
- Introduction to Machine Learning in Production
- NLP With Transformers in Python

## WORK EXPERIENCE

### Samrtblick (Berlin-Germany)

#### Data scientist

June 2019-Present

- **MarketMoverBot** (A internal chatbot for marketing and sales team): Developed a cutting-edge chatbot using Langchain (llm=gpt 3.5) and RAG technique, optimized for marketing and sales teams. Supported by retriever and memory modules.
- **OrderAnalystBot**: Developed a PostgreSQL analyst chatbot using Langchain's agent to interact with the order databases and provide answers to users' general questions.
- **Competitors analysis**: Conducted the process of evaluating the strengths and weaknesses of competitors. Leveraged NLP techniques, including variants of RAG to assist in information extraction, summarization and sentiment analysis.
- **Lead Generation**: Extracted contact information, relevant data, or potential leads from websites by using different web scraping methods.
- **Customer Analytics**: Led a comprehensive customer analytics and STP (segmentation, targeting, and positioning) framework initiative, employing hierarchical and flat clustering techniques to categorize customers into distinct groups. Utilized principal components analysis (PCA) to efficiently reduce data dimensionality.
- **Pattern counting**: Collaborated as a key member of a team focused on developing a robust algorithm for pattern recognition and similarity detection using cross-correlation.
- **Machine failure detection**: Specialized in machine health monitoring, proficiently detecting machine status and health conditions by analyzing staircase-like data sequences. Utilized advanced statistical features extraction techniques machine learning algorithms for robust pattern recognition and anomaly detection.
- **Machine data analytics**: Proficient in extracting industrial machine speed and loads by analyzing current signals, employing advanced signal processing algorithms for denoising, and extracting meaningful patterns.
- **AB testing**: A/B testing initiative to evaluate various versions of the landing page, employing statistical methods to analyze user interactions and preferences. Executed rigorous experimentation, calculating and scrutinizing p-values to determine the significance of observed differences in user responses.

### Fraunhofer-IWF, (Berlin-Germany)

Internship, Simulating heat transfer and the influence of cooling on the cutting process

Feb 2019-May 2019

- Image processing with digital microscope VHX-5000 Series
- 3D model development with NX
- FEM and CFD simulation of heat transfer between tool and chip with ANSYS Workbench

### Energy Industries Engineering & Design, (Tehran-Iran)

Project Engineer and QHSE coordinator,

Apr 2009-Sep 2015

- Data Analysis, monitoring and measurement using statistics methods such as Pareto chart and fishbone diagram,
- Monitored, analyzed & evaluated failures or qualitative problems of the project and presenting necessary mechanisms to the project management team
- Analyzed data related to the project customer satisfaction

## Education

### Technische Universität Berlin (TU Berlin)

MSc in Process, Energy and Environmental Systems Engineering (PEESE), Oct 2016 to Oct 2020

### Iran Khajeh Nasir Toosi University of Technology

MSc. in Mechanical Engineering, Sep 2005

### Iran Mazandaran University

BSc. in Mechanical Engineering, Jan 2002