Hui Zhang *Data Scientist | Machine Learning Engineer | Dr. Sc.*

Ausgebildete Neurowissenschaftlerin mit >10 Jahre Erfahrung in der Analyse großvolumiger elektrophysiologischer Daten.
Derzeit im Übergang in die Bereiche Data Science und Machine Learning Engineering.
Leidenschaftlich daran interessiert, Daten, Programmierung und analytische Problemlösungen zu nutzen, um wirkungsvolle Lösungen zu entwickeln.

Persönliche Informationen

☎: +49 1763 8678 657
✉: hzhang.davis@gmail.com
: 44892 Bochum
: HuiZhang95
: huizhang95

Tech-Stack

- **Python**:
  - Pandas
  - Numpy
  - OpenCV
  - scikit-Learn
  - Tensorflow
  - Keras
  - PyTorch
  - matplotlib
  - seaborn
  - plotly
  - Scipy
  - etc.

- **DevOps & MLOps**:
  - Git/Github
  - DVC, Dagshub
  - MLflow
  - airflow
  - Docker
  - fastAPI
  - Prometheus
  - Grafana

- **Data Analysis Tools**:
  - SQL
  - Matlab
  - R
  - SPSS

Data-Science-Projekte

Projekt 1: Klassifizierung von Rakuten-E-Commerce-Produkten

- Durchführung von Feature-Engineering für ~90k Produkte von Rakuten mit Python-Bibliotheken wie Pandas, NumPy, OpenCV, Matplotlib, Seaborn und Plotly.
- Einsatz von TensorFlow zum Training eines benutzerdefinierten rekurrenten neuronalen Netzwerks (RNN), PyTorch für das Fine-Tuning von Large Language Models (LLM; z. B. roBERTa) und Bildverarbeitungsmodellen (z. B. ResNet50) von Huggingface sowie Scikit-Learn für Feature-Fusion.
- Erreichen einer Klassifizierungsgenauigkeit von 87,25 % für neue Artikel durch die Integration von LLMs und Computer-Vision-Techniken, wodurch der 12. Platz in globalen Rankings erreicht wurde.
- Steigerung der Sucheffizienz durch die Implementierung personalisierter Produktempfehlungen.

Projekt 2: Vorhersagesystem für Verkehrsunfälle

- Durchführung von Feature-Engineering und Analyse von über 40.000 Verkehrsunfällen mit Pandas, NumPy und Scikit-learn.
- Implementierung von CI/CD-Pipelines mit Git und DVC sowie Verfolgung des Modelltrainings und der Leistung mit MLflow.
- Entwicklung einer prädiktiven API mit FastAPI, Containerisierung mit Docker und Integration von PostgreSQL zur Datenspeicherung. Überwachung der Systemleistung mit Prometheus, Grafana und Alertmanager.
- Automatisierung der Datenerfassung und Validierung mit Airflow und Evidently.
- Erreichen einer Vorhersagegenauigkeit von 82 % für Verkehrsunfälle, wodurch ein proaktives Incident-Management ermöglicht wurde.

Ausbildung

- **04/2025: Diploma in Data Science and Machine Learning Engineering**
  Paris-Sorbonne Universität (in Zusammenarbeit mit DataScientest)

- **07/2009: Doktor der Naturwissenschaften (Dr. Sc)**
  Chinesische Akademie der Wissenschaften (in Zusammenarbeit mit der University of Alberta, Kanada)

Berufserfahrung

12/2014-05/2024 **Data Scientist**
Ruhr-Universität Bochum, Deutschland

Fachkenntnisse

- Experimentelles Design
- A/B-Tests
- ML & KI
- Multivariate Statistik
- Zeitreihendaten
- Datenvorverarbeitung
- Feature-Engineering
- NLP/LLM

Kommunikation

- >20 Präsentationen auf internationalen Konferenzen
- 20 wissenschaftliche Publikationen (siehe Google Scholar) mit über 1.100 Zitierungen.

Sprachen

- Englisch
  - Verhandlungssicher
- Deutsch
  - Verhandlungssicher
- Chinesisch
  - Muttersprache

Projekt 1: Überwachung kontinuierlicher Gehirnaktivität während des Schlafs mittels A/B-Tests

- Vorverarbeitung und Feature-Engineering von Datensätzen mit ~400 Millionen Gehirnmerkmalen unter Verwendung von Python und MATLAB für multivariate Musteranalyse.
- Identifizierung der entscheidenden Rolle des Slow-Wave-Schlafs (SWS) bei der Gedächtniskonsolidierung, was Fortschritte im Verständnis von Gedächtnisdefiziten ermöglichte.

Projekt 2: Dekodierung von Verhaltensaktivitäten aus Gehirnaktivitäten

- Anwendung multivariater Musteranalyse, um verschiedene Verhaltensweisen aus über 1 Million Gehirnmerkmalen in Python und MATLAB zu unterscheiden.
- Ermöglichung der Entwicklung personalisierter Geräte für gelähmte Personen.

Projekt 3: Modellierung von Hochrisiko-Genmanifestationen für Alzheimer-Krankheit (AD) mittels A/B-Tests

- Durchführung von mehrstufigen linearen Regressions- und Mediationsanalysen an Daten von über 1.000 Personen aus dem Ruhrgebiet mit R und Python.
- Feststellung, dass kognitive Beeinträchtigungen vorwiegend bei männlichen Risikogenträgern auftreten, was den Weg für frühzeitige AD-Interventionen ebnete.

Management und Zusammenarbeit:

- Leitung eines Forschungsteams von ~5 Mitgliedern und Verwaltung einer Patientendatenbank mit über 1 Million Merkmalen pro Eintrag.
- Zusammenarbeit mit Krankenhäusern und Forschungslaboren in Deutschland, Frankreich, Spanien, den USA und China.
- Betreuung von 5 PhD-Kandidaten und über 10 Masterarbeiten sowie Lehrtätigkeit in den Bereichen experimentelles Design und Statistik.

02/2012-11/2014 **Data Scientist**

Universitätsklinikum Bonn, Deutschland

- Vorverarbeitung und Analyse von über 1 TB Patientendaten mit Excel, MATLAB und Python zur Identifizierung von Anfallsursachen bei Epilepsie.
- Mitgewirkt an der Verbesserung der automatisierten Erkennung epileptischer Spikes in vorverarbeiteten Datensätzen.

09/2009-12/2011 **Postdoktorand**

University of California, Davis, USA

- Modellierung des Lernprozesses durch die Untersuchung von über 400 Universitätsstudenten in virtuellen Realitätsszenarien unter Verwendung von Python für die Datenanalyse.
- Gewinnung von Erkenntnissen über kognitive und verhaltensbezogene Muster während des Lernens.

Hui Zhang *Data Scientist | Machine Learning Engineer | Dr. Sc.*

Trained Neuroscientist with 10+ years analysing large-scale electrophysiological data.
Currently transitioning into data science and machine learning engineering.
Passionate about leveraging data, coding, and analytical problem-solving to drive impactful solutions.

Personal Info

☎: +49 1763 8678 657
✉: hzhang.davis@gmail.com
: 44892 Bochum
: HuiZhang95
: huizhang95

Tech Stack

- **Python**:
  - Pandas
  - Numpy
  - openCV
  - scikit-learn
  - Tensorflow
  - Keras
  - pytorch
  - matplotlib
  - seaborn
  - plotly
  - Scipy
  - etc.

- **DevOps & MLOps**:
  - Git/Github
  - DVC, Dagshub
  - MLflow
  - airflow
  - Docker
  - fastAPI
  - Prometheus
  - Grafana

- **Data Analysis Tools**:
  - SQL
  - Matlab
  - R
  - SPSS

Data Science Projects

Project 1: Rakuten e-commerce product classification

- Conducted Feature engineering on ~90k products from Rakuten using python libraries such as Pandas, Numpy, OpenCV, Matplotlib, Seaborn, and Plotly.
- Utilized TensorFlow to train a custom recurrent neural network (RNN), PyTorch for fine-tuning large language models (LLM; e.g., roBERTa) and image processing models (e.g., ResNet50) from Huggingface, and Scikit-Learn for feature fusion.
- Achieved an 87.25% classification accuracy for new items by integrating LLMs and computer vision techniques, securing 12th place in global rankings.
- Enhanced search efficiency by implementing personalized product recommendations.

Project 2: Road accident prediction system

- Performed feature engineering and analysis on >40k road accidents using Pandas, NumPy, and Scikit-learn.
- Implemented CI/CD pipelines using Git and DVC, and tracked model training and performance with MLflow.
- Developed a predictive API using FastAPI, containerized with Docker, and integrated PostgreSQL for data storage. Monitored system performance using Prometheus, Grafana, and Alertmanager.
- Automated new data ingestion and validation workflows with Airflow and Evidently.
- Achieved an 82% accuracy rate in predicting road accidents, enabling proactive incident management.

Education

- **04/2025 Diploma in Data Science and Machine Learning Engineering**
  Paris-Sorbonne University (in collaboration with DataScientest)

- **07/2009 Doctor of Science (Dr. Sc)**
  Chinese Academy of Sciences (in collaboration with University of Alberta, Canada)

Professional Experience

12/2014-05/2024 **Data Scientist**
Ruhr University Bochum, Germany

Project 1: Monitoring Continuous Brain Activity During Sleep Using A/B Testing

- Preprocessed and performed feature engineering on datasets containing ~400 million brain features, utilizing Python and MATLAB for multivariate pattern analysis.

Expertise

- Experimental design
- A/B testing
- Ml & AI
- Multivariate statistics
- Time series data
- Data preprocessing
- feature engineering
- NLP

Communication

- >20 presentations at international conferences
- 20 Scientific publications (see google scholar) exceeding 1,100 citations

Languages

- English – business fluent
- German – business fluent
- Chinese – Mother tongue

- Identified the critical role of slow-wave sleep (SWS) in memory consolidation, contributing to advancements in understanding memory deficits.

Project 2: Decoding Behavioral Activities from Brain Activities

- Employed multivariate pattern analysis to differentiate various behaviors from >1 million brain features in python and Matlab
- Facilitated building personalized devices for paralyzed individuals.

Project 3: Modeling High-Risk Gene Manifestations for Alzheimer's Disease (AD) Using A/B Testing

- Conducted multilevel linear regression and mediation analysis on data from >1,000 individuals in the Ruhr area using R and Python.
- Discovered that cognitive decline is predominantly observed in male risk-gene carriers, paving the way for early AD interventions.

Management and Collaboration:

- Led a research team of ~5 members and maintained a patient database with over 1 million features per entry.
- Collaborated with hospitals and research labs across Germany, France, Spain, the USA, and China.
- Supervised 5 PhD candidates and over 10 Master's theses, while teaching courses on experimental design and statistics.

02/2012-11/2014 **Data Scientist**

University Hospital Bonn, Germany

- Preprocessed and analyzed >1TB of patient data using Excel, MATLAB, and Python to identify seizure onset patterns in epilepsy.
- Assisted in enhancing automated detection of epileptic spikes from pre-cleaned datasets

09/2009-12/2011 **Postdoctoral Scientist**

University of California, Davis, USA

- Modeled the learning process by studying over 400 university students through virtual reality scenarios, utilizing Python for data analysis.
- Gained insights into cognitive and behavioral patterns during learning.