

Sepehr Mahmoudian

SENIOR-KI-ENTWICKLER

Machine Learning • Wissensgraphen • Multimodale KI & Embodied AI

E-MAIL sepmhn@gmail.com

GITHUB github.com/sepahead

LINKEDIN linkedin.com/in/sepehr-mahmoudian

Berlin-Brandenburg • verfügbar ab Oktober 2026 • Staatsangehörigkeit: deutsch • Deutsch: gut / Englisch: fließend / Persisch: Muttersprache



PROFIL

Senior-KI-Entwickler mit über zehn Jahren Erfahrung in Forschung und Produktivbetrieb. Python-/Rust-Systeme für multimodale KI und Embodied AI: LLM-/VLM-Pipelines, Wissensgraphen, Graph RAG, VLA-Evaluation und Computer Vision.

LLMs / VLMs • Wissensgraphen • VLA-Evaluation • Modellevaluation

OPEN SOURCE • DATENPROVENIENZ • MODELLVALIDIERUNG • KI-SYSTEME

KI-ENGINEERING

10+ JAHRE

FORSCHUNG BIS PRODUKTIVEINSATZ

PYTHON • RUST • KI / ML

BERUFSERFAHRUNG

Selbstständiger KI- und ML-Entwickler & Doktorand

2020.04 - heute

Selbstständig; Promotion an der Technischen Universität Darmstadt

Berlin & Darmstadt, Deutschland

- Entwicklung quellenbelegter neuronaler Modelle, neuer PID-Analysen und von Protokollen für neuronale Steuerung; VLA-Evaluation
- Multimodale 3D-Wahrnehmung und prüfbare wissenschaftliche Visualisierung; Open-Source-Modelle und -Datensätze

Senior AI Engineer

2026.01 - 2026.03

Neural Frames UG

Berlin & Hamburg, Deutschland

- Evaluation von LLMs, VLMs und End-to-End-Pipelines; Langfuse-Tracing von Läufen und Ergebnissen
- Mitarbeit an Python-Pipelines für Modellevaluation und Inferenz

AI Engineering Resident (Hacker Room)

2025.09 - 2026.01

Merantix AI Campus

Berlin, Deutschland

- Entwicklung von Dreamweave, einem System für generative Medien, das Gaussian Splatting und Diffusionsmodelle zur räumlichen und zeitlichen Verankerung kombiniert
- Implementierung von Smart Contracts in Rust für Confidential Computing im Auftrag der NEAR Foundation
- Bereitstellung von Produktservices und Dateninfrastruktur auf Cloudflare Workers, D1, R2, Google Cloud Run und AWS EC2

Senior AI Engineer

2024.07 - 2025.07

Vision Labs UG (haftungsbeschränkt)

Berlin & Hamburg, Deutschland

- Aufbau von Workflows für Prompt-Optimierung, LLM-Evaluation und Observability mit DSPy und Langfuse
- Konzeption von Daten- und MLOps-Pipelines auf Azure und Google Cloud mit Staging, automatisierten Tests und CI/CD

Wissenschaftlicher Mitarbeiter

2021.07 - 2024.07

Freie Universität Berlin

Berlin, Deutschland

- Skalierung und Optimierung groß angelegter Simulationen neuronaler Netze in C/C++ und Python
- Entwicklung von RAG-Workflows für Fachliteratur und Evaluation der Zuverlässigkeit von KI-Agenten
- Statistische und ML-Analysen in Python; Code- und Manuskriptreviews in einem internationalen Team

Wissenschaftlicher Mitarbeiter

2019.07 - 2020.06

Georg-August-Universität Göttingen • Göttingen, Deutschland

Neuronale Netze und reproduzierbare Daten-Workflows

Wissenschaftlicher Mitarbeiter

2017.03 - 2019.07

Goethe-Universität Frankfurt • Frankfurt am Main, Deutschland • Simulation und statistische Analyse neuronaler Netzmodelle

Gastwissenschaftler

2017.03 - 2019.03

Max-Planck-Institut für Dynamik und Selbstorganisation • Göttingen, Deutschland • Statistische Methoden zur Analyse neuronaler Netze

Wissenschaftlicher Mitarbeiter

2015.02 - 2017.03

Forschungszentrum Jülich • Jülich, Deutschland • Simulationen neuronaler Netze und Experimente zum Reinforcement Learning

Softwareentwickler

2010.09 - 2012.03

Centre for Content Creation • Cyberjaya, Malaysia • Preisgekrönte iPhone- und iPad-Anwendungen sowie digitale Produkte

AKADEMISCHE AUSBILDUNG

Promotion zum Dr. rer. nat.

2020.04 - heute

Technische Universität Darmstadt • Darmstadt, Deutschland

- Dissertation in der Abschlussphase: Partial Information Decomposition (PID)
- Dissertationskern: neue mathematische PID-Analysen und informationstheoretische Repräsentationsanalyse; agentische Workflows leiten aus Computational-Neuroscience-Literatur quellenbelegte Modelle ab; Werkzeuge ermöglichen prüfbare KI-Experimente

Master's by Research in Biosciences: Neuroscience

2012 - 2014

University College London • London, Vereinigtes Königreich

Computational Neuroscience • Abschluss mit Auszeichnung (Distinction, Top 10 %, CGPA 4.0/4.0)

Bachelor of Science (Honours) in Computer Science

2009 - 2012

Anglia Ruskin University • Cambridge, Vereinigtes Königreich

Abschluss mit First Class Honours (Top 1 %, CGPA 4.0/4.0)

AUSGEWÄHLTE AUSZEICHNUNGEN & FÖRDERUNGEN

EDTH / 2025

Auszeichnung durch den Bundeswehr Cyber Innovation Hub für Computer Vision und Szenenrekonstruktion

HUMAN BRAIN PROJECT

Forschungsförderung

NEAR HORIZON

Förderung für Confidential Computing

UCL & SHEFFIELD

Im Wettbewerb vergebene Hochschulstipendien

IOS-ANWENDUNGEN

15 internationale Produktauszeichnungen

UNTERSCHRIFT

Sepehr Mahmoudian

ORT Berlin, Deutschland

DATUM 8. August 2026

ÖFFENTLICHES ENGINEERING-PORTFOLIO

GITHUB Open-Source-Projekte

github.com/sepahead

HUGGING FACE Modelle & Datensätze

huggingface.co/torusprime

AUSGEWÄHLTE PROJEKTE IM KI-ENGINEERING



engram
QUELLENBELEGTE MODELLE
Überführt neurowissenschaftliche Publikationen in quellenbelegte Modellspezifikationen und NEST-Artefakte; erfasste Modellierungsentscheidungen bleiben prüfbar.
Python Knowledge Graphs NEST
FORSCHUNGSPROTOTYP



pid-rs
INFORMATIONSZERLEGUNG
Zerlegt Information in redundante, quellenspezifische und synergistische Anteile; Methodenprovenienz und ausgewiesene Forschungsgrenzen machen Aussagen prüfbar.
Rust / Python Information Theory Statistical ML
V0.9-VORABVERSION



prisoma
PRÜFBARE VLA-EXPERIMENTE
Forschungswerkzeug für prüfbare VLA-Erfassung, Intervention und Replay; kanonische Laufprotokolle legen Aussagegrenzen und Negativbefunde offen.
Rust / Python VLA Evals Intervention / Replay
0.9-KANDIDAT



NCP
PROTOKOLL FÜR NEURONALE STEUERUNG
Projektneutrales Protokoll für Simulatoren, KI-Controller und Roboter über vier QoS-Ebenen; sicherheitsgeprüfte Aktionen und Provenienz pro Frame.
Rust Canonical JSON Zenoh
V0.8 VERÖFFENTLICHT



manwe
LUFTRAUMWAHRNEHMUNG
Forschungswerkzeug für Bilddaten, Akustik, Mehrkamergeometrie und Tracking; Modellexporte durchlaufen Artefaktidentitäts- und Numerikprüfungen.
Python PyTorch Model Export
ALPHA



crebain
RÄUMLICHES LAGEBILD
Bündelt Gaussian-Splat-Szenen, native Objekterkennung und multimodales 3D-Tracking in einer taktischen Ansicht für simulierte Drohneneinsätze.
Rust Sensor Fusion Native Inference
FORSCHUNGSPROTOTYP



melkor
INTEGRITÄT VON GAUSSIAN SPLATS
Prüft und konvertiert 3D-Gaussian-Splats deterministisch ohne GPU; legt numerische Risiken und Feldherkunft vor dem Pipeline-Einsatz offen.
C++ PLY / SPZ Asset Validation
HÄRTUNG



cobot-atlas
OFFENER ROBOTIKDATENSATZ
Zitierfähiger Robotik-Atlas: 2.023 einzigartige GLB-Meshes von Industriebauteilen bis zu Manipulationsobjekten, mit veröffentlichter Erzeugungspipeline.
Python Hugging Face 3D / VLA
VERÖFFENTLICHT + DOI



galadriel
SENSORÜBERGREIFENDE KONSISTENZ
Fail-Closed-Monitor für Sensorwidersprüche auf einer Objektspur: Innovationstests und vorzeichenbehaftete Korrelation auf deklarierten Projektionen.
Rust NIS / CUSUM Signed Correlation
FORSCHUNGSPROTOTYP



haldir
MISSIONSFREIGABEN
Prüft signierte Intents gegen Leases, vertrauenswürdigen Zustand und Richtlinien; nur ALLOW erzeugt einen neuen, vom Gateway erstellten Befehl.
Rust Authorization Signed Intents
EXPERIMENTELL

PORTFOLIO & ENTWICKLUNGSAKTIVITÄT



TECHNISCHE KENNTNISSE

KI-/ML-ENGINEERING

• KI-SYSTEME	LLMs / VLMs Generative AI AI Agents Tool Use Model Evaluation Fine-Tuning / SFT
• MACHINE LEARNING	Python PyTorch Computer Vision Multimodal AI VLA Evaluation Deep Learning
• WISSEN & RETRIEVAL	Knowledge Graphs Graph RAG RAG Embeddings Vector Search Reranking Data Provenance
• SYSTEME & MLOPS	Rust C++ FastAPI ONNX TensorRT CUDA Metal / Core ML Docker
• CLOUD & INFRASTRUKTUR	GCP AWS Azure Kubernetes Terraform GitHub Actions Linux Zenoh