

# Gabin Maxime Nguegnang

KI-/ML-Ingenieur · Applied Scientist

LLMs · RAG · KI-Agenten · Knowledge Graphs · Deep Learning · MLOps

Freising, Bayern – 85356, Deutschland | +49-176-875-103-53 | [gmaxime@nguegnang-ai.com](mailto:gmaxime@nguegnang-ai.com)

Website: [nguegnang-ai.com](http://nguegnang-ai.com) | [linkedin.com/in/maxime-nguegnang](https://linkedin.com/in/maxime-nguegnang) | [github.com/mnguegnang](https://github.com/mnguegnang)

## PROFIL

---

KI-/ML-Ingenieur und Applied Scientist mit Promotion in angewandter Mathematik. Baut Machine-Learning- und LLM-Systeme vom ersten Datenabzug bis zum laufenden Dienst und misst jedes System vor dem Einsatz. Brachte einen Random Forest in den Betrieb, den Techniker vor Ort selbst bedienen, und sicherte so 84.617 Liter Kraftstoff. Entwickelte ein Agentic-AI-System, das Geschäftsfragen an Optimierungs-Solver übergibt, mit menschlicher Freigabe ab 10.000 USD. Springer Nature veröffentlichte die Forschung dazu, wann das Training tiefer Netze konvergiert.

## AUSGEWÄHLTE ERFOLGE

---

- **Agentic-AI-System, das Geschäftsfragen an Optimierungs-Solver weiterleitet:** Entwickelte einen LangGraph-Agenten, der Supply-Chain-Fragen 10 Typen zuordnet und die Mathematik an 7 Optimierungs-Solver (OR-Tools, CVXPY) übergibt, sodass das LLM die Antwort nie selbst berechnet. Deployte das System auf Azure Kubernetes Service, mit menschlicher Freigabe ab 10.000 USD, 168 Tests und 98 % Resistenz gegen Prompt Injection.
- **Knowledge Graph, der einem KI-Agenten SAP-Geschäftskontext liefert (laufend):** Überführte die öffentlichen SALT-Vertriebstabellen von SAP (2,3 Mio. Auftragspositionen) in einen RDF/OWL-Knowledge-Graph, abfragbar per SPARQL, und verknüpfte jedes Feld mit den Definitionen von SAP, samt Quelle jeder Aussage. Nächste Stufe: ein LangGraph-Multi-Agenten-System, das diesen Kontext pro Auftrag abrufen, fehlende Felder ergänzt und an Schwellenwerten gemessen wird, die vor dem ersten Lauf feststanden.
- **84.617 Liter Kraftstoff durch einen produktiven Random Forest gesichert:** Ermittelte die Ursache der Kraftstoffverluste an Telekommunikations-Basisstationen und trainierte einen Random Forest (NSE 0,986). Lieferte das Modell als Webanwendung mit Live-Monitoring-Dashboard aus, das die Techniker selbst bedienen, und senkte die Berichtszeit von Tagen auf Sekunden.
- **Publizierte Forschung zu verlässlichem Deep-Learning-Training:** Bewies Bedingungen an die Lernrate, unter denen der Gradientenabstieg in tiefen linearen Netzen konvergiert, mit einer Schranke, die mit wachsender Tiefe nicht exponentiell schrumpft (Springer Nature, 2024). Bestätigte die Ergebnisse mit PyTorch-Training auf GPU-Clustern; bisher 5 Publikationen und 98 Zitationen.

## ANGESTREBTE POSITION

---

Angestrebt wird eine Position als AI/ML Engineer, Applied Scientist, AI Research Scientist oder Data Scientist. Am besten passt ein Team, das ein Geschäftsproblem oder eine Forschungsfrage bis zum laufenden Produkt trägt und für die berichteten Zahlen einsteht. Wohnort im Raum München, verfügbar vor Ort, hybrid oder remote, in Vollzeit.

## AUSBILDUNG

---

- Sep. 2019 – Sep. 2025*     **Promotion in angewandter Mathematik: Optimierung neuronaler Netze**  
LMU München und RWTH Aachen, Deutschland
- Aug. 2017 – Feb. 2019*     **Kooperativer M.Sc. in Industriemathematik: Machine Learning**  
African Institute for Mathematical Sciences (AIMS), Limbe, Kamerun
- Okt. 2016 – Nov. 2018*     **M.Sc. in angewandter Mathematik: Dynamische Systeme, Modellierung**  
Universität Yaunde I, Kamerun
- Okt. 2014 – Jul. 2016*     **B.Sc. mit Auszeichnung in angewandter Mathematik**  
Universität Duala, Kamerun

## BERUFSERFAHRUNG

---

- Okt. 2025 – heute*     **AI Research Consultant: LLM-Reasoning und Evaluierung agentischer KI**  
Outlier AI und Mercor, remote, Deutschland
- Analysiere Fehlermuster im mehrstufigen Reasoning von Frontier-LLMs bei MINT-Aufgaben und schreibe Korrekturen, die als RLHF-Trainingsdaten dienen.
  - Bewerte Trajektorien von KI-Agenten und die Präzision des Function Calling bei langen Aufgaben und trenne Agentenfehler von Umgebungsfehlern.
  - Formuliere die Rubrik-Kriterien, mit denen das Reward-Modell bewertet wird, und leite das Peer Review in einem verteilten Team.
- Apr. 2024 – Sep. 2025*     **Wissenschaftlicher Mitarbeiter in Forschung und Lehre**  
Ludwig-Maximilians-Universität München, Deutschland
- Entwickelte Bedingungen an die Lernrate, die die Konvergenz des stochastischen Gradientenabstiegs (SGD) garantieren, die Grundlage für verlässliches Modelltraining.
  - Vergleich 6 Ensemble-Verfahren mit 4 Baselines in einem XAI-Projekt zur Anomalieerkennung auf realen Betriebsdaten von Dieselgeneratoren und erreichte F1 0,987 mit LightGBM.
  - Verteidigte mit den Koautoren einen strittigen Beweisschritt vor der Redaktion von Springer Nature bis zur Annahme.
  - Unterrichtete Optimierung, Machine Learning, Deep Learning und Data Science im Masterstudium; betreute 1 Masterstudierenden mit.
- Sep. 2019 – Mär. 2024*     **Wissenschaftlicher Mitarbeiter in Forschung und Lehre**  
RWTH Aachen, Deutschland
- Leitete Bedingungen an die Lernrate für den Gradientenabstieg in linearen neuronalen Netzen her, die nicht exponentiell mit der Netztiefe abnehmen.
  - Validierte die Theorie und übertrug sie auf nichtlineare Architekturen, mit groß angelegtem PyTorch-Training auf GPU-Clustern.
  - Unterrichtete kontinuierliche Optimierung und Mathematik der Data Science in einer Forschungsgruppe mit 9 Nationalitäten.
- Jun. 2018 – Mai 2019*     **Praktikant Machine Learning**  
Group One Holding Company, Limbe, Kamerun
- Ermittelte per explorativer Datenanalyse die Quellen der Kraftstoffverluste an Telekommunikations-Basisstationen, trainierte einen Random Forest (NSE 0,986) und lieferte ihn mit Monitoring-Dashboard aus: 84.617 Liter Kraftstoff gesichert.
  - Automatisierte die Log-Verarbeitung (Berichtszeit von Tagen auf Sekunden) und legte mit den Wartungstechnikern fest, welche Abweichungen das System melden soll.

## PROJEKTE

---

- Agentic-ERP-Assistent für die Lieferkette** | *LangGraph, MCP, Neo4j, pgvector, OR-Tools, Azure AKS*
- Leitete 10 Anfragetypen an 7 Optimierungs-Solver hinter typisierten Pydantic-Verträgen weiter (6 FastMCP-Server). Stützte die Antworten auf hybrides Retrieval (BGE-large, pgvector, BM25) und einen Neo4j-Knowledge-Graph, abgefragt nur über eine Whitelist von Cypher-Abfragen.
  - Deployte das System aus Bicep-Vorlagen auf Azure Kubernetes Service, über eine CI/CD-Pipeline mit 6 Jobs, manuell freigegebenem Produktions-Release und 6 Smoke-Checks.
- Kontext-Engine für SAP-Kundenaufträge (laufend)** | *Python, SQL, DuckDB, RDF/OWL, SPARQL, pytest*
- Verknüpfte 4 SALT-Tabellen per SQL zu 1 Zeile pro Auftragsposition und prüfte das Ergebnis Zeile für Zeile gegen den von SAP veröffentlichten Join (2,3 Mio. Positionen).
  - Bildete die Tabellen nach dem W3C-Standard Direct Mapping auf RDF/OWL ab, ergänzte die Felddefinitionen von SAP und legte die Quelle jeder Aussage in eigenen Named Graphs ab.

### Alignment-Pipeline für Llama 3 8B | *PyTorch, Hugging Face TRL, PEFT, LoRA, DPO*

- Fine-tunte Llama 3 8B mit LoRA auf FinQA auf 1 GPU (Token-Genauigkeit von 44 % auf 80,8 %) und trainierte anschließend per DPO auf 1.800 KI-gelabelten Präferenzpaaren (Reward-Genauigkeit 97,7 %).
- Bewertete 3 Modellversionen mit einem separaten LLM-Judge (Llama 3.3 70B) auf 100 Testfragen: Die Halluzinationsrate sank von 96,0 % auf 86,7 %.

### RAG-System für wissenschaftliche Fragen | *FastAPI, FAISS, BM25, ColBERT v2, Llama 3.1 8B, Docker*

- Baute eine 4-stufige Pipeline über 47.810 Textpassagen: hybrides Retrieval, Reranking, ein Relevanzfilter, der bei schwacher Evidenz keine Antwort gibt, und Generierung mit Inline-Zitaten.
- Steigerte die Zitationspräzision auf 84 Benchmark-Fragen von 0,057 auf 0,84 und reduzierte das Deployment auf 1 Befehl.

### Coding-Agenten mit festen Engineering-Standards | *Claude Agent SDK, MCP, pytest, ruff, mypy, bandit*

- Verteilte Recherche, Coding und Review auf 3 Agenten unter 8 Quality Gates, mit Hooks, die am Sitzungsende jedes Gate erneut ausführen und ein selbst gemeldetes Bestehen blockieren, das bei erneuter Ausführung scheitert.

## AUSGEWÄHLTE PUBLIKATIONEN

---

- Nguegnang, G.M., Rauhut, H., und Terstiege, U. *Convergence of gradient descent for learning linear neural networks*. Advances in Continuous and Discrete Models, Springer Nature, 2024.
- Mulongo, G., Atemkeng, M., Ansah-Narh, T., Rockefeller, R., Nguegnang, G.M., u. a. *Anomaly detection in power generation plants using machine learning and neural networks*. Applied Artificial Intelligence, Taylor & Francis, 2020.
- Vollständige Liste, 5 Publikationen, 98 Zitationen, h-Index 3:  
[scholar.google.com/citations?user=g0C5-QAAAAJ](https://scholar.google.com/citations?user=g0C5-QAAAAJ)

## ZERTIFIKATE

---

Building GraphRAG Python MCP Tools, Neo4j GraphAcademy (Sep. 2026) | Advanced Agent Coding, Outlier AI (Jan. 2026) | Model Parallelism: Building and Deploying Large Neural Networks, NVIDIA (Okt. 2025) | Generative AI with Large Language Models, DeepLearning.AI und AWS (Okt. 2024) | Machine Learning Engineering for Production (MLOps), DeepLearning.AI (2024)

## TECHNISCHE KENNTNISSE

---

**Programmierung & Daten:** Python, PyTorch, SQL, scikit-learn, pandas, NumPy, FastAPI, Pydantic, PostgreSQL, DuckDB, pgvector, FAISS, Bash

**LLMs, RAG & KI-Agenten:** LLMs, RAG (hybrides Retrieval, Reranking), GraphRAG, KI-Agenten, Multi-Agenten-Systeme, LangGraph, LangChain, MCP, Tool Calling, Prompt Engineering, Fine-Tuning (LoRA, QLoRA), DPO, RLHF, Hugging Face (Transformers, TRL, PEFT)

**Knowledge Graphs:** Neo4j (Cypher), RDF/OWL, SPARQL

**Machine Learning & Statistik:** Deep Learning, NLP, Statistik, Random Forest, Gradient Boosting (LightGBM), Anomalieerkennung, erklärbare KI (SHAP), explorative Datenanalyse, Optimierung (OR-Tools, CVXPY)

**Evaluierung & MLOps:** LLM-Evaluierung (LLM-as-a-Judge, RAGAS, ALCE), MLflow, Docker, Kubernetes, CI/CD (GitHub Actions), Git, pytest, Azure (AKS), AWS, Google Cloud, verteiltes Training auf GPU-Clustern

## MITGLIEDSCHAFTEN & AUSZEICHNUNGEN

---

Munich Center for Machine Learning (MCML), Mitglied (Jun. 2024 – heute)

Mastercard Foundation Scholars Program Award (Aug. 2017)

## SPRACHEN

---

**Französisch:** Muttersprache (C2) | **Englisch:** verhandlungssicher (C1) | **Deutsch:** gute Kenntnisse (B2+)