

AI/ML Engineer – Transform the future of business relationships at ZopaAI

Location: Saarbrücken/Hybrid

About us: ZopaAI is a dynamic start-up developing a revolutionary software solution in the field of B2B management (business-to-business). Our software offers globally active companies an AI-based system to better manage customers and build strategic partnerships in a more targeted manner. Our goal is to change the way companies interact with each other. At ZopaAI, we take an agile approach and value creativity, teamwork, and continuous development.

As an **AI/ML Engineer**, you will contribute to the development of the AI core of our product. You will work on different stages of our machine learning systems – from data preparation and model development to deployment and continuous improvement – with a clear focus on large language models and agent-based systems.

Tasks: As an **AI/ML Engineer**, you will:

- Prepare and curate data for ML model training and evaluation (with a focus on LLMs, but not limited to that)
- Develop, fine-tune, and evaluate ML models with a focus on LLMs (e.g., Hugging Face, LoRA/QLoRA); experience with classical ML approaches (e.g., scikit-learn, PyTorch) is beneficial
- Work on prompt engineering and systematic model evaluation (benchmarks, automated test suites)
- Implement and further develop agent-based systems (tool/function calling, RAG, multi-agent orchestration)
- Take care of model serving and inference optimization (e.g., vLLM or TGI for LLMs; FastAPI/REST)
- Set up and maintain CI/CD pipelines for ML workflows (training, evaluation, deployment)

Requirements:

- Strong Python skills in the ML ecosystem (e.g., PyTorch, Hugging Face, scikit-learn, LangChain/LangGraph)
- Experience with LLMs – fine-tuning, evaluation, or deployment (e.g., LoRA/QLoRA, Ollama)
- Solid understanding of common NLP concepts
- Practical experience with prompt engineering and systematic model evaluation
- Experience building ML model APIs (e.g., FastAPI/REST) and working with containerization (Docker)
- Knowledge of CI/CD pipelines for ML workflows

- Independent and structured working style with strong problem-solving skills and a team-oriented mindset
- Experience with agent-based systems or RAG pipelines (e.g., vector databases like Qdrant/pgvector, embedding models) is a plus
- Familiarity with ML DevOps or model serving frameworks (e.g., vLLM, TGI) is beneficial
- Good German language skills (spoken and written) are a strong plus, but not mandatory

Our offer:

- Direct impact: you actively contribute to the AI core of a product that is rolled out to global industry leaders
- A steep learning curve and plenty of room to take on responsibility from day one
- The opportunity to work hands-on with state-of-the-art LLMs and agent-based systems in a real product context
- The opportunity to build your career in a fast-growing technology start-up with international ambitions
- Flexible working hours and home-office options
- Fair compensation and the potential for a long-term career path within the company
- A motivated, supportive, and international team that values your ideas and celebrates success together

If you want to be part of the growing team of ZopaAI, we look forward to your application. Please send your documents, including CV, earliest possible starting date and relevant references via email to join@zopaai.com.