

Senior AI Engineer / Architect (m/f/d)


LOCATION

Kuala Lumpur


COMPANY

ARRK Malaysia Sdn Bhd


CONTRACT TYPE

Full-time


Form of Work

Hybride

WHAT TO EXPECT

ARRK Engineering is part of the international ARRK group of companies and specializes in product development - from virtual development to prototypes and small-series production. Within our AI initiative, you will become the technical sparring partner who not only builds but also challenges architectural decisions across our growing AI product portfolio.

You will join a highly collaborative, multi-country team where technical proactiveness, outcome-oriented delivery, sound systems architecture, and the willingness to question assumptions are valued. This is a peer level role: we expect you to push back, propose alternatives, and raise the bar.

YOUR TASKS

- Act as a primary technical counterpart and architecture partner - designing, coding, and deploying a suite of custom, intelligent AI tools, and critically reviewing design decisions.
- Enhance our system's ability to manage context windows, optimize token usage, and reduce attention noise when connecting LLMs to large codebase repositories.
- Design and extend graph based code and document representations (e.g. property graphs, AST derived structures) and hybrid retrieval pipelines (dense + sparse + reranking/fusion) over vector databases.
- Translate complex algorithmic ideas - syntax parsing, semantic chunking, custom routing - into reliable, production ready software components.
- Build and rigorously evaluate prototypes for future AI initiatives, including open source small language models (SLMs), custom embeddings, and multi agent / cross validation patterns, backed by measurable retrieval and quality metrics.
- Architect multi tenant data, identity, and authorization layers that meet enterprise security and compliance requirements.
- Ensure our AI applications integrate securely into existing codebase ecosystems and developer toolchains, with both self hosted and cloud managed model serving in mind

YOUR PROFILE

- 5+ years in software engineering with substantial AI/ML product delivery, and a track record of shipping production backend systems that interface with LLMs.
- Strong conceptual grasp of how transformers process data - tokens, self-attention, embeddings, and context window limitations - and how that translates into practical engineering trade offs.
- Hands on experience with vector databases and hybrid retrieval (dense + sparse, reranking, fusion); familiarity with multimodal/visual retrieval is a plus.
- Experience with graph based representations of code or documents (property graphs, AST parsers, libraries such as Tree sitter / NetworkX) to programmatically read, prune, and structure data for LLM ingestion.
- Familiarity with both self hosted model serving (e.g. vLLM, Ollama) and cloud-managed AI services - and a clear sense of when each is appropriate, especially under data sovereignty constraints.
- Strong proficiency in Python; a systems level language (Go or Rust) is a plus where performance matters.
- Experience with major cloud provider platforms and their native managed AI services for hosting and orchestrating models.
- Awareness of security and compliance requirements in regulated environments (e.g. TISAX, EU AI Act, NIS2/DORA) and how they shape AI architecture.
- Experience with fine tuning / adapter approaches (e.g. LoRA) and structured model evaluation is a plus.

- Excellent communication skills in English; German is a plus given our multi-country setup.
- A collaborative spirit, intellectual honesty, and the confidence to challenge as well as deliver.
- A positive attitude, a collaborative spirit, and excellent communication skills.
- Passionate about building great products and continuously improving your craft

YOUR BENEFITS



Exciting projects



Attractive salary



Pleasant working
atmosphere



Hybrid work

APPLY NOW



CONTACT

Iliana Morales Reyes

career@arrk-engineering.com

+ 49 (0)89 / 31857-0

[Privacy Policy](#)

ARRK Engineering is part of the international ARRK group of companies and specialized in product development. Within the ARRK group of companies we implement product developments from virtual development to prototypes and small series production.

For reasons of simplification, only the masculine form is used in this job advertisement. However, persons of any gender and any orientation are always meant equally.

Our job offer is equally directed at severely disabled persons and persons of equal status.