



Intern with the Database AI research group!

Are you looking for an exciting internship opportunity? If so, this might be the perfect fit for you. The Database AI research group has an opening for an internship to work on the intersection of machine learning and database management systems!

Oracle

Oracle, a global provider of enterprise cloud computing, is empowering businesses of all sizes on their journey of digital transformation. Oracle Cloud provides leading-edge capabilities in software as a service, platform as a service, infrastructure as a service, and data as a service.

Oracle's application suites, platforms, and infrastructure leverage both the latest technologies and emerging ones – including artificial intelligence, machine learning, blockchain, and Internet of Things – in ways that create business differentiation and advantage for customers. Continued technological advances are always on the horizon.

Oracle Database AI Research

The group focuses on the development of AI technologies that keep Oracle at the forefront of the industry. Oracle Database researchers look for novel approaches and methodologies, often taking on projects with high risk or uncertainty, or that are difficult to tackle within the broader product-development organization. The group is focused on real-world outcomes: our researchers aim to develop technologies that will someday play a significant role in the evolution of technology and society.

Internship Details

We offer various topics based on the skills and interests of the candidate:

Agentic Enterprise AI data retrieval

First, we will design and evolve a knowledge base that captures organizational know-how about data. The focus is on abstracting patterns from diverse sources into a governed layer that encodes meaning, relationships, and usage context. Key challenges include harmonizing terminology, modeling semantics across heterogeneous assets, managing provenance and versioning, and establishing update and validation loops so the knowledge stays current as systems change.

Second, we will build higher-level agents that operationalize this knowledge. These agents will retrieve and reason over mixed data types, generate structured queries when appropriate, and adapt through user feedback. Emphasis is on modular retrieval and reasoning strategies, model routing for efficiency and quality, and mechanisms to capture corrections and preferences to incrementally improve performance. Outcomes include a maintainable knowledge representation, repeatable curation processes, prototype agents that unify retrieval and structured querying, and evaluation frameworks for accuracy, latency, and cost—developed with attention to enterprise security, privacy, and compliance.

We will also explore how to apply these techniques to domain-specific agents. The goal is to tailor knowledge representations, retrieval strategies, and query-generation behaviors to particular business domains (e.g., finance, supply chain, customer support).

Workflow and LLM inference optimizations for DB AI workloads

This internship explores how to deliver accurate, efficient AI on Oracle Database by pairing built-in DB Agents and AI operator functions with right-sized open-source CPU models. You will identify database-centric workloads where smaller models can meet accuracy requirements, benchmark them against larger baselines, and define selection criteria (latency, cost, accuracy, and resource footprint) to guide model choice in enterprise settings.

The subsequent focus will be on applying state-of-the-art LLM serving optimizations to these models. You'll research and implement techniques such as quantization, speculative decoding, optimized KV cache management, paged attention, batching, and efficient tokenization to improve throughput and reduce latency on CPU-first environments. You will also evaluate LLM routing strategies—deciding, at request time, which model (among available open-source or enterprise models) best fits the task for built-in DB Agents and AI operator functions, balancing cost-performance trade-offs while maintaining quality.

Towards Native Agent Memory in Oracle AI Database

Agentic systems are rapidly moving beyond traditional Retrieval-Augmented Generation, yet most existing memory solutions remain document-centric, fragmented, and tightly coupled to specific runtimes. We focus on building native agent memory abstractions that allow intelligent systems to persist, organize, and reason over knowledge across interactions and time.

The objective is to design and implement short- and long-term memory, conversation-level aggregates, and multi-modal knowledge graphs, backed by Oracle AI Database using vector search, graph, JSON, and access-control capabilities. The goal is to make agent memory a scalable, reusable database primitive interoperable across agent frameworks and suitable for enterprise-grade agentic applications.

Agent Technology and Open Agent Spec

We research foundational agent technologies for building autonomous, task-driven agentic systems that are deeply integrated with Oracle AI Database. Our work spans agent architectures, multi-agent composition, planning and reasoning, and safe execution over enterprise data, with the goal of enabling reliable and scalable agentic systems suitable for real-world production environments through Oracle AI Database Private Agent Factory.

A key focus of our research is enabling standardized, interoperable agent execution across frameworks and runtimes, while ensuring strong guarantees around correctness, safety, and enterprise governance. To support this, we recently open-sourced Open Agent Spec ([oracle/agent-spec](#)), a declarative specification for defining and composing AI agents. Together, these efforts aim to establish a solid foundation for building enterprise-ready autonomous agents that can reason, plan, and act securely over complex data and systems.

Semantic search on Relational Data

Enterprises store massive volumes of business-critical information in relational databases, but semantic access to this data remains limited by schema complexity and rigid query interfaces. A key research question we explore is whether we can construct relational embeddings across entities that capture the rich semantic meaning of business objects which are grounded not just in values, but in relationships, constraints, and structure inherent to relational data.

Building on advances in relational learning, which has proven effective for predictive tasks such as classification and regression, we investigate how these embeddings can be extended to support semantic search, user intent understanding, and agentic AI workflows. Relational embeddings enable semantic indexing over structured data, unlocking faster and more precise search across large-scale, multimodal relational databases. As part of this effort, you will survey and prototype relational machine learning approaches, identify and curate suitable datasets for training and evaluating relational

embedding techniques, and validate the most promising methods for scalability, robustness, and productization feasibility in enterprise-scale environments.

Required Skills

The person stepping into this role is expected to complete the internship using a wide and diverse set of skills.

- Good understanding of machine learning and AI
- Good understanding of LLMs and agentic technologies
- Experience with Python and C
- An understanding of internals of ML runtimes, LLMs, and implementation of ML kernels, is a plus
- Familiarity with SQL, system programming and an understanding of internals of database systems is a plus

Internship Facts

The duration of the internship can vary based on the candidate's constraints. The usual duration is 6 months. We pay a competitive salary. The research topics listed are informative, we are open to suggestions depending on your skills and qualifications. By sending in your application, you opt-in for processing your personal information.

If this sounds like you, we hope to meet you!

How to Apply

Please provide us with the following information to apply:

- Your CV or link to your home page containing your curriculum
- Description of your motivation and area(s) of interest
- Availability and preferred internship duration
- Preferred location

If available, please apply to this topic via the internship portal of your university. Alternatively, send an email with the aforementioned information to Vlad Haprian (vlad.haprian@oracle.com), Ingo Kossyk (ingo.kossyk@oracle.com), Damien Hilloulin (damien.hilloulin@oracle.com)

About Us

Innovation starts with inclusion at Oracle. We are committed to creating a workplace where all kinds of people can be themselves and do their best work. It's when everyone's voice is heard and valued, that we are inspired to go beyond what's been done before. That's why we need people with diverse backgrounds, beliefs, and abilities to help us create the future, and are proud to be an affirmative-action equal opportunity employer.

Oracle is an Equal Employment Opportunity Employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, national origin, sexual orientation, gender identity, disability and protected veterans status, age, or any other characteristic protected by law. Oracle will consider for employment qualified applicants with arrest and conviction records pursuant to applicable law.

Life at Oracle

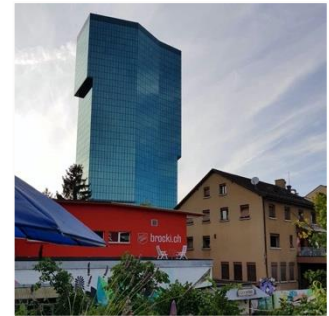


Image Caption 1. Our beautiful offices at Prime Tower in Zurich.

An Oracle career can span industries, roles, countries and cultures, giving you the opportunity to tackle new roles and challenges, while blending work and life. Oracle has thrived through 40+ years of change by innovating and operating with integrity while delivering for the top companies in almost every industry. To nurture the talent that makes this happen, we work hard to build a vibrant and inspiring workplace that celebrates diverse, hardworking teams where everyone can contribute. We take care of each other, and value giving back to the community. We have flexible work arrangements and offer benefits including generous paid parental leave and comprehensive premium medical insurance.

We will ensure that individuals with disabilities are provided reasonable accommodation to participate in the job application or interview process, to perform crucial job functions, and to receive other benefits and privileges of employment. Please contact us to request accommodation.

At Oracle, we don't just respect differences—we celebrate them. We believe that innovation starts with inclusion and to create the future we need people with diverse backgrounds, perspectives, and abilities. That's why we're committed to creating a workplace where all kinds of people can do their best work. It's when everyone's voice is heard and valued that we're inspired to go beyond what's been done before.

[More information on Oracle's stance on diversity and inclusion](#)