STANFORD UNIVERSITY

AI Applications Engineer

Stanford University

Details

Posted: 10-Sep-25

Location: Stanford, California

Type: Full-time

Categories:

Staff/Administrative

Internal Number: 107213

Job Purpose

Are you an experienced AI/GenAI engineer who loves shipping real systems? Join Stanford's Enterprise Technology team to design, implement, and support AI solutions across university use cases. In this role, you will influence strategic direction, requirements, and architecture for AI-driven information systems, incorporating new capabilities (LLMs, RAG, agentic frameworks, MLOps) to improve workflow, efficiency, and decision-making. You may serve as the technical lead for specific AI tracks and interrelated applications.

This role blends hands-on engineering with mentorship and thought leadership. You will prototype and productionize--presenting proofs of concept, demoing solutions to stakeholders, and partnering with project managers, technical managers, architects, security, infrastructure, and application teams(ServiceNow, Salesforce, Oracle Financials, etc.)

Core Duties:

- **AI/ML System Implementation & Integration:** Translate requirements into well-engineered components (pipelines, vector stores, prompt/agent logic, evaluation hooks) and implement them in partnership with the platform/architecture team.
- **Application & Agent Development:** Build and maintain LLM-based agents/services that

securely call enterprise tools (ServiceNow, Salesforce, Oracle, etc.) using approved APIs and tool-calling frameworks. Create lightweight internal SDKs/utilities where needed.

- **RAG & Search Enablement:** Configure and optimize RAG workflows (chunking, embeddings, metadata filters) and integrate with existing search/vector infrastructure--escalating architecture changes to designated architects.
- **MLOps & SDLC Practices:** Follow and improve team standards for CI/CD, testing, prompt/model versioning, and observability. Own feature delivery through dev/test/prod, coordinating with release managers.
- **Governance, Security & Compliance:** Apply established guardrails (PII redaction, policy checks, access controls). Partner with InfoSec and architects to close gaps; document decisions and risks.
- **Metrics & Reporting:** Instrument services with KPIs (latency, cost, accuracy/quality) and build lightweight dashboards. Deep BI/reporting not primary.
- **Documentation & Communication:** Write clear technical docs (APIs, workflows, runbooks), user stories, and acceptance criteria. Support and sometimes lead UAT/test activities.
- **Collaboration & Mentorship:** Facilitate working sessions with stakeholders; mentor junior engineers through code reviews and pair programming; provide concise updates and risk flags.

Education & Experience:

Bachelor's degree and eight years of relevant experience or a combination of education and relevant experience.

Required Knowledge, Skills, and Abilities:

- **Agent/Agentic Framework Experience:** Built and shipped at least one production LLM agent or agentic workflow using frameworks such as LangGraph, LangChain, CrewAI/AutoGen, Google Agent Builder/Vertex AI Agents (or equivalent). Able to explain tool selection, orchestration logic, and post-deployment support.
- **Proven Delivery:** Implemented 3+ AI/ML projects and 2+ GenAI/LLM projects in production, with operational support (monitoring, tuning, incident response). Projects should serve sizable user populations and demonstrate measurable efficiency gains.
- Strong understanding of AI/ML concepts (LLMs/transformers and classical ML) and experience designing, developing, testing, and deploying AI-driven applications.
- **Programming Expertise:** Python (primary) plus experience with Node.js/Next.js/React/TypeScript and Java; demonstrated ability to quickly learn new tools/frameworks.
- Experience with cloud AI stacks (e.g., Google Vertex AI, AWS Bedrock, Azure OpenAI) and vector/search technologies (Pinecone, Elastic/OpenSearch, FAISS, Milvus, etc.).
- Knowledge of data design/architecture, relational and NoSQL databases, and data modeling.
- Thorough understanding of SDLC, MLOps, and quality control practices.

- Ability to define/solve logical problems for highly technical applications; strong problem-solving and systematic troubleshooting skills.
- Excellent communication, listening, negotiation, and conflict resolution skills; ability to bridge functional and technical resources.

Desired Knowledge, Skills, and Abilities:

- **MLOps Tooling:** MLflow, Kubeflow, Vertex Pipelines, SageMaker Pipelines; LangSmith/ PromptLayer/Weights & Biases.
- **Open Source Savvy:** Experience working with, customizing, and improving open-source solutions; comfortable contributing fixes/features upstream.
- **Rapid Tech Adoption:** Demonstrated ability to pick up a new technology/framework quickly and deliver production value with it.
- **GenAI Frameworks:** LangChain, LlamaIndex, DSPy, Haystack, LangGraph, Agent Engine, Google ADK, AWS AgentCore, CrewAI/AutoGen. Security & Governance: Implementing AI guardrails, red-teaming, and policy enforcement frameworks.
- **Enterprise Integrations:** ServiceNow, Salesforce, Oracle Financials, or others.
- **UI Development:** React/Next.js/Tailwind for internal tools.
- **Prompt engineering at scale:** Structured prompts (JSON/function-calling), templates, version control; automated/offline & online evals (rubrics, hallucination/bias checks, A/B tests, golden sets).
- **Parameter-efficient fine-tuning** (LoRA/QLoRA/adapters), supervised instruction tuning; hosting open-weight models (Llama/Mistral/Qwen) with vLLM/TGI/Ollama.
- **Safety/guardrails frameworks** (Guardrails.ai, NeMo Guardrails, Azure/AWS safety filters) and jailbreak/drift detection.
- **Hybrid search & reranking** (BM25+dense, Cohere/Voyage/Jina rerankers), synthetic data generation, provenance/watermarking.
- **Telemetry & governance:** prompt/model drift monitoring, policy-as-code, audit logging, red-teaming playbooks.

Certifications and Licenses:

- Required: One of (or equivalent experience with): Google/AWS/Azure ML/AI certifications or strong demonstrable portfolio of production AI systems.

Physical Requirements*:

- Constantly perform desk-based computer tasks.
- Frequently sit, grasp lightly/fine manipulation.

- Occasionally stand/walk, writing by hand.
- Rarely use a telephone, lift/carry/push/pull objects that weigh up to 10 pounds.

**Consistent with its obligations under the law, the University will provide reasonable accommodations to applicants and employees with disabilities. Applicants requiring a reasonable accommodation for any part of the application or hiring process should contact Stanford University Human Resources by submitting a contact form.*

Working Conditions:

- May work extended hours, evenings, and weekends.

Work Standards:

- Interpersonal Skills: Demonstrates the ability to work well with Stanford colleagues and clients and with external organizations.
- Promote Culture of Safety: Demonstrates commitment to personal responsibility and value for safety; communicates safety concerns; uses and promotes safe behaviors based on training and lessons learned.
- Subject to and expected to stay in sync with all applicable University policies and procedures, including but not limited to the personnel policies and other policies found in Stanford's Administrative Guide, http://adminguide.stanford.edu.

The expected pay range for this position is $169,728 to $190,000 per annum.

Stanford University provides pay ranges representing its good faith estimate of what the university reasonably expects to pay for a position. The pay offered to a selected candidate will be determined based on factors such as (but not limited to) the scope and responsibilities of the position, the qualifications of the selected candidate, departmental budget availability, internal equity, geographic location and external market pay for comparable jobs.

At Stanford University, base pay represents only one aspect of the comprehensive rewards package. The Cardinal at Work website (https://cardinalatwork.stanford.edu/bene...) provides detailed information on Stanford's extensive range of benefits and rewards offered to employees. Specifics about the rewards package for this position may be discussed during the hiring process.

Why Stanford is for You:
Stanford University has revolutionized the way we live and enrich the world. Supporting this mission is our diverse and dedicated 17,000 staff. We seek talent driven to impact the future of our legacy. Our culture and unique perks empower you with:

- **Freedom to grow.** We offer career development programs, tuition reimbursement, or audit a course. Join a TedTalk, film screening, or listen to a renowned author or global leader speak.
- **A caring culture.** We provide superb retirement plans, generous time-off, and family care resources.
- **A healthier you.** Climb our rock wall, or choose from hundreds of health or fitness classes at our world-class exercise facilities. We also provide excellent health care benefits.
- **Discovery and fun.** Stroll through historic sculptures, trails, and museums.
- **Enviable resources.** Enjoy free commuter programs, ridesharing incentives, discounts, and more.
- **Redwood City.** Our new Stanford Redwood City campus, opened in 2019, will be the workplace for approximately 2,700 staff, including University IT, whose jobs are important to supporting the University's mission. The campus will offer amenities such as onsite cafes and a dining pavilion, a high-end fitness facility with an outdoor pool, and a childcare center for Stanford families.

The job duties listed are typical examples of work performed by positions in this job classification and are not designed to contain or be interpreted as a comprehensive inventory of all duties, tasks, and responsibilities. Specific duties and responsibilities may vary depending on department or program needs without changing the general nature and scope of the job or level of responsibility. Employees may also perform other duties as assigned.

Stanford is an equal employment opportunity and affirmative action employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, protected veteran status, or any other characteristic protected by law.

STANFORD
UNIVERSITY

About Stanford University

Why work at Stanford?Stanford University has changed the world, over and over again.We are one of Silicon Valley's largest employers - and also one of the most unique. Our mission is to educate future leaders and promote interdisciplinary, world-class research and teaching. This passion makes Stanford an intensely creative, rewarding, and challenging place to work. At the same time, our traditions of respect and collaboration sustain a humane, supportive environment in which to pursue your life and your career.At Stanford you'll work with bright, diverse, dedicated people. You'll find encouragement to learn and grow. You'll enjoy excellent benefits and an outstanding environment. How will it change you?

https://main.hercjobs.org/jobs/21655537/ai-applications-engineer