English Speaking STAFF AI ENGINEER

募集職種

人材紹介会社
株式会社PROGRE

採用企業名
Enterprise-size Fintech Company

求人ID
1619525

業種
インターネット・Webサービス

会社の種類
大手企業 (300名を超える従業員数)

外国人の割合
外国人 半数

雇用形態
正社員

勤務地
東京都 23区, 港区

給与
1100万円 ~ 2000万円

勤務時間
9:30 - 18:30

更新日
2026年11月26日 02:00

応募必要条件

職務経験
6年以上

キャリアレベル
中途経験者レベル

英語レベル
ビジネス会話レベル (英語使用比率: 常時英語)

日本語レベル
無し

最終学歴
専門学校卒

現在のビザ
日本での就労許可が必要です

募集要項

You will lead the design and delivery of production-grade AI agents for customer-facing products. You will tackle complex technical challenges at the intersection of agent engineering, backend systems, and infrastructure at scale, turning prototypes into reliable, secure, and cost-efficient systems used by real customers.

This is a hands-on, high-impact individual-contributor role. You will shape technical direction for building, evaluating, deploying, and operating agentic systems, including durable runtimes for long-running tasks, while continuing to write code

and raising engineering standards across teams.

Key Responsibilities

- Agentic System Design: Architect, build, and scale multi-agent systems and LLM-powered services for customer-facing products, from prototype through production.
- Agent Engineering: Design reliable tool use, function calling, memory, multi-turn interactions, protocols, and orchestration frameworks. Establish guardrails, evaluation methods, and safe fallback behavior.
- Backend and Infrastructure at Scale: Own backend services, APIs, and agent infrastructure, ensuring high availability, low latency, secure secrets and credential handling, and horizontal scalability.
- Evaluation and Benchmarking: Build end-to-end evaluation harnesses, including representative task sets, fixed inputs and reference outcomes, repeatable runners, and scoring for task completion, correctness, latency, cost, and human-intervention rate. Establish LLM-as-a-judge and offline/online evaluation, integrate regression gates into CI, and develop a benchmark suite that Product and QA teams can use as an acceptance gate.
- Durable, Long-Running Agent Execution: Design runtimes for tasks lasting ten minutes or more, including task lifecycle management, isolated execution sandboxes, artifact generation and retrieval, retries, back-off, checkpointing, resume, and self-recovery.
- Reliability and Cost Optimization: Identify bottlenecks and optimize latency, throughput, token and compute costs, and reliability. Instrument observability across the agent lifecycle.
- Safety and Governance: Apply security and governance practices appropriate for customer-facing systems, including input validation, content filtering, PII handling, human-in-the-loop escalation, and risk-based sampling.
- Cross-Functional Technical Leadership and Mentorship: Work with Product, Business, Backend, Frontend, Infrastructure/SRE, and QA teams to translate business goals into agent architecture, align standards across teams, drive cross-team system decisions, and mentor engineers in agent development, LLM integration, and evaluation methodology.

スキル・資格

Required Skills and Experience

- 7+ years of professional software engineering experience, with recent hands-on delivery experience; this is not a purely managerial role.
- Deep backend engineering expertise, including designing, building, and operating high-throughput, low-latency production systems and secure APIs.
- Strong infrastructure skills with AWS and/or Azure, Docker, Kubernetes, Terraform, and CI/CD, including production operations and troubleshooting.
- Proven hands-on experience building AI agents or agentic orchestration for real products using modern frameworks, especially the Claude Agent SDK (agentic loop, tool use, sessions, sandboxed execution, and Skills). Comparable depth with LangGraph or similar frameworks is relevant, along with sound judgment about when to use a single agentic loop versus a multi-stage pipeline.
- Strong Python or TypeScript skills for production services, such as FastAPI/FastMCP, asynchronous request handling, dependency and lifecycle management, and ASGI/Node runtimes.
- Solid understanding of MCP, REST APIs, and GraphQL for tool integration and agent-to-agent communication.
- Demonstrated experience building agent evaluation from scratch: designing task sets and scoring rubrics, running controlled baseline-versus-variant experiments, and using results to guide architecture decisions.
- Experience with LLM tracing and observability (OpenTelemetry/OpenLLMetry, Langfuse, or equivalent), as well as performance and load testing of production services.
- Experience building or operating durable, long-running execution infrastructure, including job/task lifecycle management, isolated sandboxes (containers, microVMs, or managed sandbox services), artifact storage, and failure recovery.
- Strong fundamentals in computer science, including data structures, algorithms, software design, and engineering best practices.

Preferred Skills and Experience

- Experience shipping AI features in customer-facing or B2B SaaS products under real security and compliance constraints.
- Experience with guardrails, safety, and governance frameworks for LLM/agent systems, such as OWASP-style threat modeling for agents.
- Experience optimizing LLM costs through context compression, model routing, use of non-frontier models, or gateways.
- Fluent use of AI-assisted development tools such as Claude Code, Cursor, GitHub Copilot, or Codex, with good judgment about when to delegate to AI and when to verify.

会社説明