

AI Engineer

Remote (US) | Remote | Full-time | Department: Product Engineering

| | | | |
|--------------|-------------------------|--------------|--------------------|
| Reports to | Engineering Manager, AI | Team | Support Automation |
| Level | Mid-Level | Travel | Up to 5% |
| Posting date | February 24, 2026 | Start window | Next 4-8 weeks |

About the company

Paperplane Support builds B2B SaaS for customer support automation. Our teams ship software frequently, measure outcomes, and invest in reliability and security. We build AI-assisted workflows for agents: summarization, suggested replies, and knowledge retrieval with strong safety and privacy controls.

Role summary

Build AI capabilities across our products, spanning classical ML and modern generative AI. You will develop prototypes, harden them into production services, and partner with teams to deliver customer value.

What you will do

- Develop AI features using ML and/or LLMs, from prototype to production
- Implement data preparation, prompt/agent logic, evaluation harnesses, and deployment pipelines
- Design APIs and services to integrate AI into existing products securely and reliably
- Create automated tests and evaluation metrics; monitor quality, latency, and cost
- Collaborate with product, design, and security to define safe usage and guardrails
- Document decisions, runbooks, and operational procedures for AI services

What we are looking for

Required

- 3+ years of software or ML engineering experience
- Strong Python skills; experience shipping backend services (REST/gRPC) and writing tests
- Familiarity with ML or LLM concepts: embeddings, retrieval, fine-tuning, and evaluation basics
- Experience working with data stores and pipelines (SQL, event logs, warehouses)
- Comfortable troubleshooting production issues across data, model, and code

Preferred

- Experience with RAG systems, vector databases, and offline/online evaluation

- Experience with model governance, security reviews, and privacy-by-design
- Familiarity with GPU workloads or model serving frameworks

Tech stack (typical)

Python, FastAPI, PyTorch, Vector DB, Docker/Kubernetes, CI/CD, Observability; Vector DB, RAG evaluation harness

Success looks like (first 90 days)

- Deliver an end-to-end model improvement with an agreed evaluation plan and measurable lift
- Harden a training or inference pipeline (tests, monitoring, runbooks) and reduce operational toil
- Partner with product to define a rollout plan; participate in an A/B test or controlled launch
- Mentor a junior engineer through at least one project milestone

Compensation & benefits

Base salary range: \$150,000 - \$195,000 per year. Includes eligibility for bonus depending on level. Final offer depends on experience, skills, and location.

- Competitive salary and annual bonus or equity (role dependent)
- Medical, dental, and vision coverage
- 401(k) or local retirement plan with employer match (where applicable)
- Flexible PTO and paid holidays
- Learning budget for courses, conferences, and books
- Home-office/remote-work stipend (where applicable)

Interview process

- Intro call with Talent Partner (20-30 min)
- Technical screen (coding + ML/system fundamentals)
- Onsite/virtual loop (2-4 interviews: ML, systems, product, collaboration)
- Final conversation with hiring manager

We are an equal opportunity employer. All qualified applicants will receive consideration without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, veteran status, or any other legally protected status.

If you require an accommodation during the application or interview process, please let us know. We will work with you to meet your needs.