No Japanese Required AI Solution Architect (Speech Product)

English speaking role

募集職種

人材紹介会社
株式会社PROGRE

採用企業名
AI company

求人ID
1561703

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, 新宿区

給与
700万円 ~ 1300万円

ボーナス
給与： ボーナス込み

勤務時間
10:00〜19:00

休日・休暇
完全週休二日制 所定休日：土・日・祝日 休暇：年次有給休暇、夏季休暇（3日）、年末年始休暇（12月31日〜1月3日）、慶弔休暇

更新日
2026年01月21日 01:00

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
大学卒： 学士号

現在のビザ
日本での就労許可が必要です

募集要項

Role Summary

You will join the Solution Architect team as a Speech Product specialist. You will lead backend development for custom solutions that solve customer business challenges using AI Speech product and voice AI technologies. We expect you to handle the entire process from system design and implementation based on customer requirements to integration with existing systems and operational maintenance.

Why You'll Love This Role

- **Cutting-edge voice AI technology challenges:**
  - Work on next-generation voice dialogue systems combining ASR/TTS/LLM technologies

- **Real-time processing expertise:**
  - Build highly marketable expertise through designing and operating low-latency voice streaming infrastructure using WebRTC/Twilio

- **Frontline of customer problem-solving:**
  - Engage directly with customers across diverse industries and experience the impact of solving their challenges with technology

- **Socially impactful product development:**
  - Work on high-impact projects directly contributing to corporate productivity improvements, such as call center efficiency and sales support

- **Collaboration with diverse talent:**
  - Grow in a stimulating environment working alongside talented members from diverse backgrounds including AI engineers, SREs, PdMs, Sales, and CS

- **Balance of personal and organizational growth:**
  - Thrive in a dynamic field that balances individual growth with organizational development while adapting flexibly to rapid change

Responsibilities

- **Voice AI Product Development**
  - Design and develop custom solutions for customers utilizing JAPAN AI Speech product
  - Backend implementation of real-time voice processing systems (ASR/TTS/dialogue)
  - Build low-latency voice streaming infrastructure using WebRTC / Twilio / SIP
  - Implement voice dialogue systems utilizing large language models (LLMs)
  - Build voice data collection, preprocessing, and metadata management pipelines

- **System Integration & API Connectivity**
  - Design and develop API integrations with existing systems (CRM/CTI/business systems)
  - Design distributed systems using microservices architecture
  - Design and implement REST/gRPC/GraphQL APIs
  - Implement webhook design and event-driven architecture

- **Infrastructure & Operations**
  - Containerization and Kubernetes operations on GCP
  - Database design and query optimization (Postgres/Redis/Elasticsearch, etc.)
  - Design and operate streaming infrastructure (Kafka / Pub/Sub, etc.) and batch ETL
  - Build and improve CI/CD pipelines (GitHub Actions / ArgoCD, etc.)
  - Build monitoring and observability systems (Prometheus/Grafana/ELK/Jaeger)

- **Machine Learning & AI Integration**
  - Integrate and deploy speech recognition and speech synthesis models to production environments
  - Optimize inference API performance and reduce latency (Triton / TorchServe / ONNX, etc.)
  - Build and operate MLOps pipelines

- **Customer Engagement & Project Management**
  - Requirements definition, technical explanations, and contract development with customers
  - PoC design, implementation, and evaluation
  - Technical leadership and project management
  - Maintenance and operational support for developed solutions
  - Implementation and operation based on security requirements (voice data encryption, masking, retention policies)

Work Style

Hybrid work model: 3 days in the office, 2 days remote.

スキル・資格

Minimum Qualifications

- Bachelor's degree in Computer Science, Software Engineering, Artificial Intelligence, Machine Learning, Mathematics, Physics, or related fields, or equivalent practical experience
- 3+ years of application development experience using Python
- Experience with requirements definition, technical explanations, and contract development with customers
- Practical experience with real-time communication (WebRTC / WebSocket / gRPC, etc.) or streaming processing (Kafka / Pub/Sub, etc.)
- 3+ years of experience with cloud services (GCP preferred, AWS/Azure acceptable) and container technologies (Docker, Kubernetes)
- Experience designing and operating RDB/NoSQL (Postgres, Redis, etc.) and object storage (GCS/S3, etc.)
- Experience working at startups or growth-stage companies, or equivalent environments
  Knowledge of and passion for AI-related technologies

Preferred Qualifications

- Practical experience with voice products (ASR/TTS/voice signal processing/speaker separation, etc.)
- Production experience with model inference (Triton, TorchServe, ONNX Runtime, TensorRT, etc.)
- CTI integration experience with Twilio / SIP / Asterisk
- Experience operating large-scale data infrastructure (BigQuery, Redshift, Snowflake, etc.) or data lakes
- Implementation experience with dialogue systems combining LLMs (GPT series, Gemini, etc.) and RAG
- System development experience utilizing natural language processing or machine learning
- Machine learning model development experience using Python, TensorFlow, PyTorch
- Experience with chatbot, speech recognition systems, or AI agent development
- Full-stack application development experience (TypeScript/React)
- Experience launching new businesses/products from the 0→1 phase
- Customer negotiation or project management experience
- Bilingual (Japanese/English)

会社説明