

Ryan Stephens

AI Engineer

712-310-3735 @ ryan.stephens15@gmail.com <https://ryan-stephens.dev/>
Omaha, NE



SUMMARY

I'm an AI Engineer with 10+ years of software engineering experience, specializing in production LLM applications and MLOps infrastructure. Recently built a medical data extraction platform with systematic evaluation and testing, along with a self-hosted AI infrastructure template for rapid prompt experimentation and deployment. Experienced with AWS Bedrock, LangChain/LangGraph for building conversational agents, and retrieval-augmented generation for grounding AI responses in real data.

Strong foundation in Python development, cloud infrastructure, and docker-based deployments, with hands-on experience implementing complete observability systems including tracing, quality scoring, and regression testing for AI applications. Leverage AI-augmented development tools to accelerate learning and iteration. Bring a quality-first engineering mindset from years of full-stack development, now applied to building reliable AI products that work in production. Remote-first engineer who thrives on solving complex problems independently while collaborating with cross-functional teams to ship features that deliver real business value.

PROFESSIONAL EXPERIENCE

Senior Software Engineer  KBX Technology Solutions Koch	08/2021 - 11/2025 Remote
Contributed to a major legacy-to-modern platform transformation, developing TMS 2.0 as a scalable SaaS application using Angular, TypeScript, .NET/C#, and cloud-native technologies (Kubernetes, Kafka, ArgoCD, Docker). Built multiple business-critical tools including bid analysis, order management, and competitive pricing systems while managing weekly production deployments via CI/CD pipelines (GitHub Actions). Technical work spanned full-stack development including microservices architecture, gRPC APIs, PostgreSQL/MongoDB databases, state management (NGRX), and responsive UI development with Angular Material, while maintaining high code quality through comprehensive unit testing (Karma, Jasmine) and contributing to technical decision-making. <ul style="list-style-type: none">Developed full-stack features for TMS 2.0 platform using Angular, TypeScript, .NET/C#, and microservices architecture, contributing to legacy-to-modern transformation supporting scalable SaaS operations.Built business-critical applications including bid analysis tool, order management system, and competitive pricing engine, delivering end-to-end solutions from requirements gathering through production deployment.Managed weekly production and pre-production deployments using CI/CD pipelines (GitHub Actions), coordinating releases, host configurations, and system status updates to ensure reliable delivery.Implemented cloud-native infrastructure using Kubernetes, Docker, Kafka, and ArgoCD, enabling scalable microservices deployment and real-time data processing capabilities.Developed responsive, mobile-first UI components using Angular Material, NGRX state management, and RxJS, ensuring consistent user experience across devices.Built RESTful and gRPC APIs with .NET/C# backend services, integrating with PostgreSQL and MongoDB databases to support high-volume transaction processing.Maintained code quality through comprehensive unit testing (Karma, Jasmine), code reviews, and adherence to software engineering best practices across greenfield and legacy applications.Enhanced system observability by implementing DataDog monitoring across production environments, improving reliability and enabling proactive issue detection.Utilized LaunchDarkly feature flags to enable controlled rollouts and A/B testing, reducing deployment risk and enabling rapid iteration.Collaborated with cross-functional teams including product, QA, and operations to deliver features aligned with business objectives and customer requirements.Mentored junior developers on best practices including TypeScript patterns, Angular architecture.	

SKILLS

LLM & Gen AI

AWS Bedrock · OpenAI API · LangChain · LangGraph · Langsmith · LLM Fine-Tuning · RAG · Streaming Responses (SSE) · Prompt Engineering/Versioning · Multi-turn Conversational AI · AI Agents · Vector Embeddings & Semantic Search

MLOps & Production AI

Systematic Evals · AI Tracing · Quality Grading · Kubernetes · Docker · MLflow · Model Registry · Backtests · Regression Testing · A/B Testing · Cost Optimization & Token Management · Model Performance Monitoring · CI/CD · Dataset Management & Curation

Python & ML Frameworks

Python · FastAPI · PyTorch · Tensorflow · Pydantic · Hugging Face Transformers · Celery · PyPDF2 · NumPy · Pandas · Pytest

Databases & Vector Search

PostgreSQL · pgvector · DynamoDB · Redis · FAISS · SQLAlchemy · Prisma ORM · Pinecone · Weaviate

Cloud & Infrastructure

AWS (Lambda · S3 · Bedrock · Sagemaker) · Azure (DevOps · ML Studio) · Docker · Kubernetes · AWS CDK · Terraform · Loki · Grafana · Prometheus · Apache Kafka · CI/CD Deployments

Full-Stack Development

React · Next.js · Angular · TypeScript · Javascript · C# · .NET · Node.js · NestJS · TailwindCSS · Vite · Responsive Design

PROFESSIONAL EXPERIENCE

 Advanced Programmer Analyst	01/2017 - 07/2021
QAT Global	Remote

Progressed from QA Engineer to Advanced Programmer Analyst across 9+ years, contributing to HIPAA-compliant healthcare applications including a risk assessment platform for healthcare insurance and Vitalistics, a home sleep apnea testing system. Developed full-stack solutions using .NET/C#, built automated test suites with Selenium, and established secure AWS cloud desktop infrastructure enabling HIPAA-compliant remote access for medical professionals analyzing patient reports.

- Developed HIPAA-compliant healthcare applications using .NET/C# including a risk assessment platform for healthcare insurance and Vitalistics, a home sleep apnea testing system connecting patients with medical providers.
- Architected and deployed secure AWS cloud desktop infrastructure (Amazon WorkSpaces) enabling HIPAA-compliant remote access for physicians to analyze patient sleep study reports while maintaining data security and privacy compliance.
- Migrated enterprise .NET application from AWS to Azure as the sole developer, reducing annual hosting and infrastructure costs by over \$12,000 while maintaining zero downtime during transition.
- Built automated test suites using Selenium for cross-browser compatibility testing, ensuring application quality and standardization across healthcare risk assessment platform.
- Developed REST APIs using ASP.NET integrated with SQL Server databases, creating secure backend services consumed by Angular front-end applications via HttpClient.
- Optimized Angular application performance through lazy loading, route optimization, and module structuring, while implementing static site generation using Scully for faster page loads.
- Established Angular coding standards and best practices enforced through CI/CD pipeline scripts (Azure DevOps), automatically blocking non-compliant code commits to maintain code quality.
- Coordinated with offshore development teams to deliver full-stack features, managing requirements and code through Azure DevOps with daily Scrum ceremonies.
- Maintained Windows Forms applications integrated with existing .NET systems, ensuring HIPAA compliance and secure handling of protected health information (PHI).
- Built and managed Azure CI/CD pipelines for automated testing, build validation, and deployment across development, staging, and production environments.

 Intern	05/2012 - 12/2016
QAT Global	Omaha, NE

Began career as QA Engineering Intern, building foundational skills in software testing, test automation, and healthcare application development. Created comprehensive manual test suites for HIPAA-compliant risk assessment platform, then automated testing workflows using Selenium, establishing early expertise that led to progression into full-stack development roles.

- Created comprehensive test cases and designed testing workflows for HIPAA-compliant healthcare risk assessment application, documenting process flows and application logic to guide manual QA testing.
- Performed manual functional and regression testing across web applications, identifying bugs and working with development team to ensure quality standards before production releases.
- Provided IT infrastructure support including employee workstation setup and onboarding, network administration, server management, and general office security maintenance.
- Assisted accounting department with operational tasks and collected marketing research data to support business development initiatives.
- Documented complex application workflows and system integrations, creating visual process maps that improved team understanding of SaaS insurance platform architecture.

SKILLS

Development & Testing

Git · GitHub Actions · SOLID · DRY ·

Agile/Scrum · E2E Testing ·

API Documentation · Postman · Rest API ·

gRPC API · DataDog

MY PROJECTS

AI Infrastructure Template - Self-Hosted MLOps Platform

Domain-agnostic AI infrastructure template for rapid AI product development. Compositional prompt architecture, evaluation framework with 6 matcher types, response caching, and complete observability stack.

Tech Stack: Python, FastAPI, Docker, PostgreSQL (pgvector), Redis, MinIO, Prometheus, Grafana, Loki, AWS Bedrock, React, TypeScript

 <https://ai.ryan-stephens.dev>

MedExtract - Medical Data Extraction & AI Agent Platform

Production-grade AI system combining automated medical document extraction with a LangGraph ReAct agent featuring 8 specialized tools. Implemented systematic evaluation framework, complete observability (custom traces + LangSmith integration), LLM-based quality scoring, and regression testing with backtests.

Key Features: RAG implementation, dual interaction types (extraction + conversational), field-level precision tracking, automated quality grading

SpeakEasy - AI Integrated Desktop Application

Built a production Electron desktop application enabling developers to interact with AI agents using voice commands while maintaining file context awareness. Implemented speech-to-text for hands-free coding assistance, file injection for context-aware responses, and real-time AI communication

Full-stack application using Next.js, TypeScript, PostgreSQL (Prisma ORM), Docker containerization, Stripe payment integration, and Playwright for E2E testing. Deployed as a windows desktop app with authentication and subscription management.

 [https://speakeeasydev.com/](https://speakeasydev.com/)

PROFESSIONAL EXPERIENCE



IT Technician	08/2014 - 08/2015
Iowa State University	Ames, IA

Progressed from IT support technician to head technician, managing workstation maintenance, security, and software imaging for Iowa State University IT infrastructure. Provided technical support to students and faculty while taking on increasing responsibility for system configuration, hardware troubleshooting, and standardized software deployment.

- Provided technical support for over 3,000 computer systems campus-wide, troubleshooting hardware and software issues across Windows and Mac operating systems.
- Managed workstation deployment processes, including system imaging, disk encryption, security configuration, and standardized software installation for new and retiring equipment.
- Advanced to the head technician role, overseeing system maintenance standards, security protocols, and software image management for university IT infrastructure.
- Delivered responsive customer service to students, faculty, and staff while managing high-volume support tickets and maintaining SLA compliance.

EDUCATION



Bachelor's in Management Information Systems	2012 - 2016
Iowa State University of Science and Technology	Ames, IA

GPA | **3.3** / 4.0

REFERENCES

KBX - Stewart Sweet, Team Lead

KBX - Sunil Khandwala, Senior SWE

KBX - Ryan Clouse, Senior SWE

WHAT I'M PROUD OF

★ My Decision Making Skills

Strong at evaluating technology tradeoffs and selecting the right tools for each challenge. Whether choosing between REST vs. gRPC for API design, evaluating state management solutions (NGRX vs. alternatives), or deciding on deployment strategies (Kubernetes vs. serverless), I research thoroughly, weigh trade-offs, and make pragmatic decisions that balance technical excellence with business needs. This analytical approach extends to leveraging AI-augmented development tools (Windsurf, Cursor, Claude) to accelerate learning new frameworks while maintaining code quality.

★ My Focus & Persistence

Known for driving complex projects to completion, even when facing technical challenges or ambiguity. When migrating QAT's enterprise .NET application from AWS to Azure solo, I worked through integration issues, infrastructure differences, and deployment complexities to deliver successfully with zero downtime. When managing weekly TMS 2.0 production deployments at KBX, I coordinated releases across microservices, handled rollback scenarios, and ensured reliable delivery despite tight deadlines. I don't give up when encountering obstacles - I research solutions, leverage AI tools to accelerate problem-solving, and persist until I've delivered a quality result.