

To Do List 8.25.25

- ✓ Create app using prompt eng in code
 - completed on 8/26!

Ideas:

Customer Interview automation

- Jobs to be done:
 - Identify person to interview
 - Add context to why this person should be interviewed
 - What questions to ask this person
 - Capture notes from interview
 - Send a follow up thank you email
 - Store interview notes, contact data, interview date
- Architecture
 - URL scraper e.g. for linkedin
 - LLM
 - Notetaker
 - Calendar
 -
 - CRM
 - Email
- Functions:
 - Calendar Function
 - Communication Function
 - Notetaker Function
 - Research Function(LLM)
 - Identifies current domain and previous domains
 - identifies current role and company
 - identifies previous roles and company
 - categorizes the companies by domain e.g. blockchain, financial services, manufacturing, software, technology infrastructure, consumer software, enterprise software, lending, banking, insurance, consulting services, education, government, etc

- Uses company products/services, known company customers, company domain(s), specific jobs to be done for role, jobs to be done for their team, to develop questions to ask AS a startup founder conducting user research, market research, and specifically looking to identify pain points for the customer, their business, and their specific team.
- Analyst Function(LLM)
- Assistant Function

Todo:

- ☒ Module 3 "additional resources"
- ☒ [DeepLearning.ai](#) courses on RAG & Finetuning
- ☐ Work on AI project idea assistant or another idea?

To D

OpenAI Key

Open AI

test key 2 - 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

Building YouTube Summarizer

To use the UV library:

1. navigate to correct folder
2. activate virtual environment in the correct folder
3. install requirements.txt with listed libraries and necessary information. must run using 'uv pip install -r requirements.txt'

4. run 'jupyter lab' with `uv run jupyter lab`

Import and Export Research

<https://nepc.gov.ng/importer/nigeria-product/sesame/>

https://www.tendata.com/blogs/tradedata/5157.html?utm_source=chatgpt.com

<https://www.shopify.com/enterprise/blog/largest-b2b-marketplaces>

<https://comtradeplus.un.org/>

Free/Public Sources

These give you country-level trade flows, sometimes down to 6-digit HS codes (but usually not importer names):

- **UN Comtrade**
 - Free, official UN data.
 - Search by **Reporter = United States, Partner = Nigeria**, and filter by HS code (e.g., 1801 for cocoa beans, 3102 for urea).
 - Returns trade value (USD), weight, and trends.
 - No company names — just commodity flows.
- **U.S. Census Bureau – USA Trade Online**
 - Free, but requires registration.
 - Official U.S. import/export stats by HS code, down to port level.
 - Lets you track Nigeria → U.S. imports in detail.
 - Again: commodity values, not company names.
- **USITC DataWeb**
 - Free, but needs signup.
 - Gives detailed U.S. trade data at the HS-10 level.
 - More granular than Census, still no company names.
- **TradingEconomics or OEC**
 - Good for a quick snapshot, charts, and country comparisons.
 - Less detailed than UN Comtrade/USITC.

Paid / Commercial Platforms

If you want to see **which companies** are importing (e.g., IMAD International, Olam Americas, Monroe Energy, JR Simplot), you'll need a customs-data provider:

- **Panjiva (S&P Global)** – Very widely used, subscription-based. You can search by HS code, country, and see **importer/exporter company names**.
- **ImportGenius** – Lets you search by HS code, shipment records, importer/exporter names.
- **Volza** – Cheaper than Panjiva, shows shipment counts and some buyer names for Nigeria → U.S. trade.
- **Datamyne (Descartes)** – Another big customs-data provider.

Practical Workflow

2151. Use **UN Comtrade / USITC / Census Trade** to confirm which HS codes and how much value comes into the U.S. from Nigeria.

- Example: HS-1801 (cocoa beans), HS-3102 (urea), HS-0714 (yams).

2152. Then plug those HS codes into **Panjiva / ImportGenius / Volza** to pull **importer-level detail**.

2153. Sort results by **value or shipment count** to identify the largest U.S. buyers.

In the period from Nov 2023 to Oct 2024, 46 importers were active with [Imad Int LLC](#), [IMAD INTERNATIONAL LLC](#) and [Taola Bisis Trade African Food In](#) accounting for 61% of United States's in nigeria total Foodstuff imports.

Imad Int LLC is the leading Foodstuff importer in United States buying from Nigeria, constituting 28% of the total with 5 shipments. Following closely is IMAD INTERNATIONAL LLC, with a 22% share of the total, equivalent to 4 shipments. Taola Bisis Trade African Food In occupies the third spot, making up 11% of the total with 2 shipments.

Source: https://www.volza.com/p/foodstuff/buyers/buyers-in-united-states/coo-nigeria/?utm_source=chatgpt.com

Petroleum and Oil Import Data: https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?f=M&n=PET&s=MTTIMUSNI1&utm_source=chatgpt.com

Key Takeaways:

- Oilseed, cocoa, and tea collectively accounted for 59% of the total agricultural exports from Nigeria to the United States.

- Oilseed meal and cake constituted the United States' largest agricultural import from Nigeria, valued at \$29.9 million.
- Total US agricultural imports from Nigeria in 2024 amounted to \$114.7 million.
- Cocoa beans represented 19% (\$21.8 million) of Nigeria's agricultural exports to the US.

Source: https://intelpoint.co/insights/oilseed-meal-cake-top-the-list-of-nigerias-agricultural-exports-to-the-united-states-occupying-26-of-the-total-exports-in-2024/?utm_source=chatgpt.com

To Do List 8.21.25

Find code repo to clone and create automated task.

Keep it simple!

Watch the week 2 video

conducting market research on reddit for trends and consumer insights

conducting market research on reddit for trends and consumer insights

Here's how to effectively use Reddit for market research:

1. Define Your Research Question:

- Clearly identify what information you want to gather and why.
- For example, are you looking for feedback on a new product concept, identifying pain points with existing products, or understanding consumer sentiment around a particular topic?

2. Identify Relevant Subreddits:

- Search for subreddits related to your industry, product category, or target audience.
- For example, a clothing brand might monitor subreddits like r/fashion or r/streetwear.

3. Utilize Reddit's Search Function and Third-Party Tools:

- Use Reddit's search function to find relevant keywords, phrases, and discussions.
- Consider using third-party tools like [Brandwatchor](#) [Sprout Social](#) for more advanced analysis and sentiment tracking.
- [OneClick](#) can also help summarize Reddit conversations and identify key themes and emotional sentiments.

4. Analyze Trends and Sentiments:

- Pay attention to recurring themes, popular topics, and common pain points.
- Analyze upvotes and downvotes on trending conversations to gauge user sentiment.
- Look for unanswered questions or areas where users are seeking specific information.

5. Engage with the Community (When Appropriate):

- If relevant to your research, consider engaging in discussions or asking questions in a respectful and authentic way.
- Be transparent about your affiliation with the brand and avoid overly promotional language.
- Focus on community engagement rather than direct selling.

6. Turn Insights into Actionable Strategies:

- Use the gathered insights to refine your product development, marketing campaigns, and overall business strategy.
- For example, if you discover a common pain point, you can address it in your product or marketing messaging.

Key Considerations:

- **Be mindful of Reddit's culture:** Avoid overly promotional or spammy content.

- **Focus on community building and engagement:** Don't just focus on selling your product.
- **Be transparent about your intentions:** Let users know you're conducting market research.
- **Be patient and persistent:** It takes time to build trust and gather meaningful insights.

On-Premise OCR/AI System for Invoice Data Extraction (No Cloud / No Commercial APIs)

Summary (PDF version of this job post is attached)

We are looking for an experienced AI/ML developer to build a production-grade, on-premise OCR/AI system for automated extraction of data from invoices (PDF, JPG, PNG).

The system must run fully on our infrastructure, export structured JSON data, and support adaptive learning based on user corrections.

Core workflow and key features:

1. Input: o Purchase document files (e.g., PDF, JPEG/PNG images).
2. Preprocessing: o Conversion of document to image format (if PDF). o Image preparation: noise removal, contrast alignment, page segmentation. o Extraction of individual elements (e.g., sections, tables).
3. Text Recognition (OCR/AI): o Utilization of a robust OCR/AI model for high-accuracy text extraction from images. o Precise localization and identification of all key document elements:
 - Document Number
 - Issue Date
 - Sale Date
 - Buyer (Name, VAT ID/NIP, Address)
 - Seller (Name, VAT ID/NIP, Address)
 - Invoice Line Items (Name, Quantity, Unit Price, Net Value, Tax Amount, Gross Value, Tax Rate)
 - Summary (Total Net, VAT, Total Gross)
 - Seller's Bank Account Number
 - Order/Purchase Order Number (if present on the invoice).
4. Data Analysis and Structuring: o Parsing of recognized text into a structured format (e.g., JSON). o Accurate mapping of all extracted data to the corresponding fields in the defined structure.
5. Data Validation and Correction: o Automatic verification of data correctness (e.g., dates in the correct format, sums calculated correctly). o Use of machine learning models (e.g., classification, named entity recognition) to improve recognition, especially when OCR/AI results are uncertain.

6. Output Export: o Structured document in JSON format, containing all extracted and validated data.
7. Machine Learning and Adaptation: o The system must be pre-trained with the necessary knowledge to accurately recognize files from the outset. o When a user makes a manual correction to the extracted data, the system automatically saves these corrections as training data in a versioned dataset; these corrections are submitted to the system via the API.
 - ▮ Collection of corrections: The system shall provide a mechanism for collecting user corrections to extracted invoice fields, submitted via a dedicated API endpoint (/submit-corrections). All submitted corrections shall be recorded and stored locally.
 - ▮ Manual retraining trigger: Model retraining may be initiated manually by the Client via a dedicated API endpoint (/trigger-retrain) or by using a provided script (e.g., retrain.py). The system shall not perform retraining automatically without explicit action by the Client.
 - ▮ Retraining process: New model versions shall be trained locally, using only the corrections collected by the Client. The retraining process will rely exclusively on open-source tools and shall not require any external services (e.g., cloud solutions, commercial APIs, etc.).
 - ▮ Versioning and rollback: Each new model version created through retraining shall be saved as a separate artifact, along with metadata (e.g., retraining date, corrections used, base model identifier). The system shall enable the Client to revert to any previous model version (rollback) in the event of degraded extraction quality.
 - ▮ Transparency and audit trail: The system shall maintain a register of all corrections and model version history. All operations shall be performed entirely locally, with no data sent to any external servers.

Summary of key requirements for the delivered system:

- On-premise Deployment: The solution must run entirely on our local infrastructure.
- API-Driven: the system must provide comprehensive API endpoints for:
 - o Processing an input file and returning structured JSON output.
 - o Submitting user corrections to feed the adaptive learning loop.
 - o Manually triggering retraining of the model based on submitted corrections.
 - o Retrieving system status and metadata (e.g., current model version, last retraining date, document processing statistics).
- Multi-language Support: The system must process invoices in multiple languages, mostly Polish. The OCR/AI model must be highly effective for various languages and common invoice layouts typical for European markets.
- Accuracy Requirement: The system must achieve at least 95% field-level extraction accuracy on the provided invoices in order for milestones to be accepted.

Preferred technology stack We strongly prefer the use of proven, open-source, locally deployable tools, such as:

- OCR & Preprocessing: Tesseract, PaddleOCR, TrOCR, pdf2image, OpenCV, Pillow
- Document Layout & Field Detection: Donut, LayoutParser, DocTR
- NER and Data Structuring: spaCy (with custom models), Hugging Face

Transformers (e.g., XLM-RoBERTa, mBERT) • Backend & Deployment: FastAPI, Python, Docker • Data Handling & Storage: SQLite/PostgreSQL (for corrections, logs, versioning) • ML Frameworks for Retraining: PyTorch, scikit-learn, Hugging Face Other fully open-source and on-premise alternatives may be proposed, provided they meet all project requirements (on-prem deployment, adaptive learning, no commercial/cloud APIs).

Project milestones Milestone 1 – Data Extraction Demo & Accuracy Test (\$1,500) • The contractor delivers the code and installation instructions so that we can run the pipeline on our own server. • The pipeline must include a temporary UI (e.g., Streamlit/Flask) that allows us to upload invoices and view extracted data. • At this stage, we only test the data extraction accuracy. The goal is to confirm at least 95% field-level accuracy on our provided invoices. • No API or feedback loop is required yet – only extraction and verification.

Milestone 2 – Full System Delivery (\$1,500) • Delivery of the complete production-ready system, including: o API endpoints, o correction feedback mechanism and manual retraining with versioning/rollback, o final technical documentation and installation guide. • Code provided as Dockerized service or equivalent, fully deployable on our server. • Temporary UI remains available for acceptance testing purposes only. Additional Provision We expect the contractor to provide active support during deployment if we encounter difficulties installing the system ourselves based on the provided instructions.

Training data provided: • ~1,000 invoices available for model training. • For ~100 invoices, we provide ground truth annotations to support supervised evaluation and fine-tuning.

Final provisions:

1. Transfer of Code Ownership Upon full payment for the project, all intellectual property rights, including source code, models, documentation, and any other materials related to the project, will be transferred to the client.
2. Confidentiality of Invoice Data The contractor agrees to maintain strict confidentiality regarding all invoices, data, and other materials provided for the project, and not to disclose them to any third party or use them for any purpose other than the execution of this project.
3. Project Acceptance and Completion The project will be considered complete only after the solution has been successfully deployed on the client's infrastructure, all technical documentation and detailed installation/migration instructions have been delivered, and the solution has been formally accepted by the client.

How to apply: In your proposal, please include:

1. Your proposed technology stack for this project (open-source, on-premise only).
2. Short description of similar projects you have delivered (preferably invoice OCR/NER systems).

3. Confirmation that you can meet the requirements: on-premise, fine-tuning enabled, no commercial APIs.

This project is not a proof-of-concept. We are seeking a contractor who can deliver a robust, complete solution. We are flexible regarding the exact technology stack — you are welcome to propose the tools and models you find most suitable — but:

- The solution must be 100% on-premise (no cloud dependencies, no commercial APIs).
- It must allow fine-tuning and retraining locally, based on user corrections.
- It must not be based only on prompt engineering with large LLMs. We value flexibility before the contract is signed, but after the contract is agreed, we expect full adherence to the chosen approach. Please propose the stack you actually plan to implement.

TrueStoriesAI Technical Overview and Migration

TrueStoriesAI Technical Overview

Project Description

TrueStoriesAI is an enterprise-grade AI-powered customer insight platform that transforms qualitative research data (interviews, focus groups, reports) into actionable business intelligence. The platform uses advanced natural language processing and retrieval-augmented generation (RAG) to extract deep insights about customer drivers, barriers, segments, journeys, and brand perceptions.

Core Technology Stack

Backend:

- Python 3.11 with Flask web framework
- SQLite database (development)
- Pinecone vector database for semantic search
- Voyage AI for embeddings and neural reranking
- Anthropic Claude Opus 4 for analysis and synthesis
- JWT-based authentication
- RESTful API architecture

Frontend:

- React 18 with TypeScript
- Vite build tooling
- Tailwind CSS for styling
- Real-time token management

- Multi-format export capabilities

Infrastructure:

- Currently runs on local development servers
- Three separate services: Main app (8081), Frontend (8080), Admin Portal (8082)
- File-based storage for documents
- Environment-based configuration

Key Features

1. **Advanced Retrieval System:** Combines vector search, neural reranking, and dynamic relevance thresholds
2. **V3 Data Archaeology Architecture:** Organizes evidence for Claude to analyze rather than attempting analysis
3. **Multi-Stage Reporting:** TMAPS workflow for comprehensive TV/media analysis
4. **Usage Tracking:** Comprehensive API usage monitoring for billing
5. **Analysis Context Management:** Cumulative knowledge building across analyses
6. **Admin Portal:** Full administrative interface for platform management
7. **Professional Export:** Multi-format export (Markdown, HTML, DOCX, PDF)

Current Architecture Strengths

- Modular design with clear separation of concerns
- Comprehensive test coverage
- Well-documented codebase
- Feature flag system for progressive rollout
- Robust error handling and logging
- Clean API design with consistent patterns

Current Limitations

- SQLite database not suitable for concurrent users
- No horizontal scaling capability
- File-based document storage lacks redundancy
- Basic authentication without 2FA
- No caching layer for performance
- Manual deployment process
- Limited monitoring and alerting
- No backup/disaster recovery
- Development-grade security throughout

Production Migration Brief for Development Team

Project Overview

You will be migrating TrueStoriesAI from a development environment to a production-ready, scalable cloud infrastructure. The application is currently functional but requires significant infrastructure upgrades, security hardening, and operational tooling to support enterprise customers.

Current State

- **Database:** SQLite (single file at /truestoriesai.db)
- **File Storage:** Local filesystem for uploaded documents
- **Authentication:** Basic JWT without 2FA or session management
- **Deployment:** Manual process with shell scripts
- **Infrastructure:** Three separate services running on different ports
- **Security:** Development-grade with hardcoded secrets in .env files

Required Production Changes

1. Database Migration (Critical Priority)

Current: SQLite with 30+ tables, ~100MB of data **Target:** PostgreSQL 15+ with connection pooling

Tasks:

- Set up PostgreSQL with read replicas for scaling
- Implement connection pooling (pgBouncer or similar)
- Create migration scripts from SQLite to PostgreSQL
- Update all database queries for PostgreSQL compatibility
- Implement proper indexes based on query patterns
- Set up automated backups with point-in-time recovery
- Configure monitoring and slow query logging

Key Tables to Review:

- `api_usage` - High write volume, needs partitioning by date
- `projects` - Core business entity, needs careful migration
- `users` - Requires encryption for sensitive data
- Vector data remains in Pinecone (no changes needed)

2. Authentication & Security (Critical Priority)

Current: Basic JWT with email/password **Target:** Enterprise-grade authentication with 2FA

Tasks:

- Implement proper 2FA (TOTP/SMS/Email options)
- Add OAuth2/SAML support for enterprise SSO
- Implement session management with Redis
- Add rate limiting and brute force protection
- Implement API key rotation mechanisms
- Add audit logging for all authentication events

- Set up IP allowlisting for enterprise accounts
- Implement CORS properly for production domains
- Add Content Security Policy headers

3. File Storage & CDN (High Priority)

Current: Local filesystem storage **Target:** Cloud object storage with CDN

Tasks:

- Migrate to S3 (or equivalent) for document storage
- Implement secure signed URLs for document access
- Set up CloudFront (or equivalent) CDN
- Add virus scanning for uploaded files
- Implement document encryption at rest
- Create migration script for existing documents
- Add file type validation and size limits
- Implement multi-region replication for availability

4. Infrastructure & Deployment (High Priority)

Current: Manual deployment with shell scripts **Target:** Containerized with orchestration

Tasks:

- Create Docker containers for all services
- Set up Kubernetes cluster (EKS/GKE/AKS)
- Implement blue-green deployment strategy
- Create Helm charts for configuration
- Set up GitHub Actions CI/CD pipeline
- Implement proper health checks and readiness probes
- Configure auto-scaling based on load
- Set up service mesh for inter-service communication
- Implement proper secret management (AWS Secrets Manager/Vault)

5. Monitoring & Observability (High Priority)

Current: Basic Python logging to files **Target:** Full observability stack

Tasks:

- Implement structured logging with correlation IDs
- Set up Prometheus + Grafana for metrics
- Add APM (DataDog/New Relic) for performance monitoring
- Implement distributed tracing (Jaeger/Zipkin)
- Set up error tracking (Sentry)
- Create custom dashboards for business metrics
- Implement SLI/SLO monitoring
- Set up PagerDuty integration for alerts

6. API & Performance (Medium Priority)

Current: Synchronous Flask API **Target:** Async with caching and queuing

Tasks:

- Migrate to FastAPI or add async support
- Implement Redis caching layer
- Add API response caching where appropriate
- Implement job queue (Celery/RQ) for long-running tasks
- Add WebSocket support for real-time updates
- Implement GraphQL API alongside REST
- Add API versioning strategy
- Implement request/response compression

7. Security Hardening (Critical Priority)

Tasks:

- Security audit of all dependencies
- Implement WAF (Web Application Firewall)
- Add DDoS protection
- Implement encryption in transit (TLS 1.3)
- Regular penetration testing schedule
- Implement security headers (HSTS, etc.)
- Add API request signing for sensitive operations
- Implement data masking for PII in logs
- Set up security scanning in CI/CD pipeline

8. Compliance & Data Privacy (High Priority)

Tasks:

- Implement GDPR compliance (right to deletion, data export)
- Add SOC 2 compliance logging
- Implement data retention policies
- Add privacy policy acceptance tracking
- Implement data residency controls
- Add cookie consent management
- Create data processing agreements template
- Implement audit trail for all data access

9. Operational Excellence (Medium Priority)

Tasks:

- Create runbooks for common operations
- Implement chaos engineering practices
- Set up staging environment matching production
- Create load testing suite
- Implement feature flags service (LaunchDarkly/similar)

- Add A/B testing framework
- Create operational dashboards
- Implement SRE practices and error budgets

10. Cost Optimization

Tasks:

- Implement auto-scaling policies
- Add spot instance usage where appropriate
- Optimize database queries and indexes
- Implement caching strategies
- Add cost monitoring and alerting
- Review and optimize AI API usage
- Implement data lifecycle policies
- Add resource tagging for cost allocation

Technical Contacts & Resources

External Services to Maintain:

- Pinecone API (vector database)
- Voyage AI API (embeddings)
- Anthropic API (Claude)

Critical Configuration:

- Pinecone indexes use empty namespace (must maintain)
- JWT tokens use HS256 (should upgrade to RS256)
- Database has WAL mode enabled (replicate in PostgreSQL)

Success Criteria

1. 99.9% uptime SLA capability
2. 200ms API response time (p95)
3. Support for 1000+ concurrent users
4. Automated deployment completion
5. RTO 1 hour, RPO 5 minutes
6. SOC 2 Type II compliance ready
7. Zero-downtime deployments
8. Comprehensive monitoring coverage

Risk Mitigation

1. **Data Migration:** Create rollback plan, test extensively
2. **Performance:** Load test at 2x expected capacity
3. **Security:** External security audit before go-live
4. **Downtime:** Implement gradual rollout with fallback

5. Cost: Set up budget alerts and scaling limits

Offshore AI Automation & Web Development Partner (Agency or Freelancer)

Needs to hire 3 Freelancers Summary About Us: We're a fast-growing AI automation agency based in Atlanta, GA, specializing in helping businesses streamline operations and improve efficiency. We are looking for a highly skilled and dedicated offshore partner.

Preferably a (whitelabel) agency, but will consider a top-tier individual to handle the technical fulfillment of client projects. This partnership will allow us to focus on client acquisition and communication while you lead the technical execution and build-out.

The Role: You will be our technical expert, responsible for the end-to-end development and implementation of a wide range of AI and web automation projects. This includes everything from building complex workflows to deploying web applications and managing domain and hosting settings. A proven track record and a deep understanding of the specified tools are essential for success.

Key Responsibilities:

- Designing, building, and deploying AI-powered automations using a variety of platforms.
- Developing and managing chatbots with tools like Voiceflow, Synthflow, Intercom etc.
- Building and deploying applications using modern platforms such as Bolt.new and Netlify.
- Integrating APIs and third-party services, including OpenAI, Airtable, and various CRMs.
- Providing domain and hosting management for our other ventures, which involves working with registrars and hosts like GoDaddy and Bluehost.
- Ensuring all solutions are robust, scalable, and well-documented.
- Collaborating with our team to meet project deadlines and client expectations.

Required Skills and Expertise:

- Extensive experience with AI automation platforms: Mostly n8n and also Make.com and Zapier.
- Chatbot development proficiency: Voiceflow, Synthflow, Intercom etc..
- Database and data management skills: Airtable, Supabase.

- Web development and deployment expertise: Strong knowledge of Bolt.new and Netlify for building and hosting applications.
- Domain & Host Management: Experience with platforms like GoDaddy, Bluehost, and general DNS/hosting configuration.
- Demonstrated portfolio of past projects and case studies.
- Strong problem-solving skills and meticulous attention to detail.
- Excellent communication skills and the ability to work independently.

How to Apply: (IGNORING THIS = WE IGNORE YOUR APPLICATION)

- Submit your proposal with a detailed overview of your experience. Highlight your expertise with the specific tools and platforms mentioned above.
- A portfolio or case studies showcasing your work in AI automation, chatbot development, and web deployment are a significant plus.
- Please specify whether you are an agency or an individual freelancer.
- Start your application with the word "Global".

We are seeking a long-term, mutually beneficial partnership. We look forward to hearing from you!

AI Agents + Automations Specialist - Project Manager

Summary

About the Role

We are seeking a RAG Agent and Implementation Specialist to lead the design and implementation of AI agents and AI automations that redefine business workflows. You will be a key driver of innovation, developing scalable solutions that optimize user experiences while ensuring compliance and seamless system integration.

Your Key Responsibilities

Develop AI Solutions — Create and deploy AI agents for websites, CRM systems, and more.

Train and Optimize AI Agents — Enhance AI agents to handle client queries effectively with built-in human oversight.

System Integration — Integrate AI solutions with business platforms to support business operations.

End-to-End Project Management — Lead projects from ideation to deployment and ongoing support.

Collaborate with Stakeholders — Work with leadership and client teams to define and execute AI-driven strategies.

Monitor, Improve, and Comply — Establish best practices for AI performance, troubleshooting, and data privacy.

What You Bring

AI Expertise — Experience building AI agents for digital platforms.

Problem-Solving Skills — Strong analytical ability to turn business needs into technical solutions.

Collaboration and Leadership — Exceptional interpersonal skills and project ownership.

AI Voice Tech Knowledge (Preferred) — Familiarity with AI Voice workflows is a plus.

What We Offer

A leadership role at the forefront of AI business transformation.

Competitive compensation, bonuses, and a comprehensive benefits package if moving full time.

Access to cutting-edge AI technologies.

A collaborative, innovative, and supportive environment.

Flexible working arrangements.

AI Agent Expert Wanted – Automation & Process Optimization for Growing Agency

Summary

Description:

Our agency is scaling fast, and we want to take our internal operations to the next level.

We're looking for an AI Agent Expert who can set up intelligent workflows, build AI Agents, and automate as well as optimize processes such as lead research and qualification using tools like Clay, Relevance AI, Apollo.io, and Ocean.io.

Your Responsibilities:

- Build and implement AI Agents to automate sales and marketing processes (e.g., lead research, data enrichment, email sequences)
- Integrate tools such as Clay, Relevance AI, Apollo.io, and Ocean.io into seamless, high-performance workflows
- Develop smart data pipelines for research, prioritization, and outreach
- Analyze existing processes, identify bottlenecks, and create AI-driven optimizations
- Continuously monitor and improve AI Agents for maximum efficiency and results

Who We're Looking For:

- Proven experience setting up AI Agents and workflow automation
- Strong expertise with tools like Clay, Relevance AI, Apollo.io, Ocean.io (or similar platforms)
- Excellent analytical skills with a systems-thinking approach
- Experience optimizing marketing and sales operations
- Independent, proactive, and results-driven working style with a focus on delivering measurable impact

Key Point:

Our clients are always at the center of what we do. The automations you build will not only increase our internal efficiency but will also directly contribute to helping our clients achieve faster, better results.

Why Work With Us:

- Long-term collaboration with growing responsibility
- Opportunity to design processes from the ground up
- Access to cutting-edge AI tools and solutions
- Close collaboration with an ambitious leadership team

\$20.00 - \$55.00

Build AI Sales Assistant for CRM Lookups, Email Drafting, and SOP Search (OpenAI / LangChain / RAG)

Summary Title: Build AI Assistant for CRM Lookups, Email Drafting, and SOP Search (OpenAI / LangChain / Repfabric)

Job Description: We're a U.S.-based electronics manufacturer rep firm using Repfabric CRM, and we're seeking a developer to build a modular AI assistant to support three internal workflows:

🧠 What We Want Built (Phase 1 MVP):

1. Sales Lookup Agent Query customer and NBO data via natural language

Sample prompts:

"What are my open NBOs in Utah?"

"Summarize activity for [account name]"

2. Email Generator Agent Draft personalized sales emails based on CRM context and templates

Sample prompts:

"Write a follow-up email after product intro"

"Send quarterly check-in to top Colorado accounts"

3. Internal SOP Agent Search internal process and policy docs (from SharePoint/OneDrive)

Sample prompts:

"How do I log a principal visit?"

"Where is our Q3 sales deck?"

⚙️ Tech Stack (Preferred): OpenAI GPT-4 / GPT-4o

LangChain or LlamaIndex

Pinecone, Weaviate, or Chroma

Microsoft Graph API (optional, for SharePoint integration)

Local or web-based UI (chat, dashboard, or Teams bot)

📁 Data Availability: We are not doing full data cleanup before this project — you'll work from:

Structured exports (CSV) from Repfabric

Real email examples for template tuning

A curated set of internal docs from SharePoint/OneDrive

You'll help identify data gaps or cleanup priorities during the pilot.

📅 Timeline: We're targeting 4–6 weeks for the MVP. We're open to phased delivery per module.

💰 Budget: Fixed price: \$12,000 (negotiable based on experience/scope). Open to milestone payments or hourly if appropriate.

✅ What to Include in Your Proposal: Examples of past GPT/LangChain/RAG projects

Your suggested architecture or process

Time and budget estimate

Any experience with Microsoft 365 or CRM data

You will be asked to answer the following questions when submitting a proposal:

Have you built an AI agent using LangChain or LlamaIndex? Please describe the use case, your role, and the outcome.

Have you integrated GPT-4 with structured data (e.g., CSV, CRM exports) or business documents for Q&A or search? Briefly explain your approach and tools use

What would be your preferred method for indexing and querying documents from SharePoint or OneDrive? Describe whether you'd use APIs, manual ingestion, or another method.

Can you propose a high-level timeline and fixed price (or milestone-based breakdown) for this 3-part MVP? Include number of weeks, delivery checkpoints, and cost estimate.

Bonus: Do you have experience working with CRMs that are not natively API-rich (like Repfabric or industry-specific tools)? If yes, how did you work around limitations (e.g., using exports, scraping, etc.)?

Expert Mentor in LLM Fine-Tuning & Multimodal AI (Interactive Collaboration + Career Prep)

Summary I'm an early-career ML practitioner with solid self-taught experience in Python, machine learning, and fine-tuning large language models (LLMs). I've previously worked with a mentor for a year and have a good grasp of key concepts like: Hugging Face Transformers & Datasets PEFT / LoRA / QLoRA for efficient fine-tuning Vision-Language Models (e.g., Qwen2-VL, BLIP, Flamingo) Using tools like trl, peft, bitsandbytes, and accelerate Python scripting for model training and evaluation

However, I now need close, interactive mentorship from a professional who:

Has industry or research experience in LLMs or multimodal AI Can collaborate with me on real mini-projects Will review and refactor my code Can guide me through best practices in: Modular code structuring Training/debugging pipelines Data preprocessing (for instruction-style and multimodal tasks) Model evaluation and deployment readiness work closely with me, teach me interactively Develop a 3-month learning roadmap to become

job-ready Learn to debug and improve my own fine-tuning scripts Build a solid portfolio of 1–2 small projects (e.g., fine-tuning Qwen2-VL or LLaVA on a specific domain task)

Understand code standards, version control, best practices, and how to make my work reproducible and modular Grow into a role that can contribute to an AI/ML team in fine-tuning, RLHF, or multimodal model applications

Technical Areas to Cover (Some or All): LLM and VLM Fine-Tuning (LoRA, QLoRA, PEFT) HuggingFace Transformers, Datasets, TRL Image-text/multimodal finetuning (e.g., Qwen2-VL, LLaVA, BLIP-2) Evaluation: BLEU, ROUGE, bounding box detection accuracy

Debugging memory, multi-GPU configs, and accelerate workflows Model & dataset structuring best practices

Less than 30 hrs/week Hourly 1 to 3 months Duration Expert I am willing to pay higher rates for the most experienced freelancers \$10.00

•

\$40.00

Hourly

Applied ML/NLP Engineer — SaaS MVP

Summary

Description:

We're building a new SaaS platform and need an Applied ML/NLP Engineer to design and implement the intelligence layer. This role involves analyzing text data, extracting structured insights, generating recommendations, and integrating those into a SaaS workflow.

You'll work closely with a Full-Stack Engineer who handles the APIs, dashboard, and integrations.

Responsibilities:

Develop models/pipelines to parse and analyze text data.

Extract entities, relationships, and sentiment from structured/unstructured text.

Implement a scoring system based on presence, context, and frequency of entities.

Build a content analysis/recommendation engine that identifies gaps and suggests improvements.

Create guardrails to ensure suggestions follow specific style/brand constraints.

Collaborate on prompt templates and LLM-based enrichment tasks.

Work with backend engineer to integrate outputs into SaaS workflows and CMS APIs.

Tech Stack (preferred experience):

NLP/ML frameworks: spaCy, Hugging Face Transformers.

LLM APIs: OpenAI, Anthropic, Google Gemini.

Embeddings for semantic similarity.

Experience with text classification, NER, keyword/entity extraction, or semantic scoring.

Ideal Candidate:

4+ years applied NLP/ML engineering.

Strong background in practical NLP pipelines.

Comfortable using LLMs for enrichment tasks with constraints.

Bonus: experience with content optimization or recommendation systems.

Scope:

Initial MVP (~45 days): Entity/mention parsing, scoring, content suggestion pipeline.

Future work: Expanded enrichment features, deeper semantic scoring, automation improvements.

Generative AI Specialist (Avatars & Media) Summary

Summary

Generative AI Specialist (Avatars & Media) (Stable Diffusion + ComfyUI)

Role Summary:

Our team is seeking an experienced AI content creator specialized in Stable Diffusion and ComfyUI. You'll be responsible for developing and maintaining digital human models—ensuring consistent visual and audio identity (face, skin tone, voice, accent) across multiple 25–60 second ad videos. These models will serve as long-term assets, so strong organization, reusability, and collaboration are key.

The Opportunity:

- 1) Scalability & Brand Consistency: Your creations will form the core of our digital human asset library, essential for coherent multi-ad campaigns.
- 2) Workflow Optimization: By working closely with our internal team—and especially the current role-holder—you'll help refine, document, and scale AI workflow processes across our organization.
- 3) Creative-Technical Hybrid: This isn't just execution; it's strategic content creation powered by AI, balancing technical precision with storytelling.

Key Responsibilities

Avatar Creation

- 1) Build and refine 5+ older male and 5+ female digital human models.
- 2) Preserve consistency in facial features, voice, accent, and skin tone across all generated outputs.
- 3) Test across varied environments/backgrounds (maintaining model integrity).

Output Delivery & Organization

- 1) Deliver each model's content into clearly structured folders for easy reuse and scalability.
- 2) Ensure file naming, versioning, and documentation are user-friendly for non-technical team members.
- 3) Collaboration & Process Building.
- 4) Work closely with our internal creative and technical team member.
- 5) Co-develop best practices, refine existing workflows, document process improvements.

Quality Assurance & Iteration

- 1) Ensure output quality aligns with brand standards and storytelling goals.
- 2) Iterate based on feedback—customer-facing stakeholders may review for realism, consistency, and tone.

Required Skills & Tools

- 1) ComfyUI — Deep understanding of node-based generative pipelines, workflow saving/reuse.
 - 2) Stable Diffusion — Text-to-image and fine-tuning experience, including image-to-image and consistency workflows.
 - 3) Veo 3 (or equivalent AV generation tools).
- AI Character Modeling — Strong knowledge of maintaining facial, tonal, and vocal consistency across outputs.
- 4) Generative Image-to-Video Tools — Prior experience with synthetic media tools/techniques.
 - 5) File & Workflow Organization — Folders, naming conventions, version control.

Nice-to-Have (Preferred):

- 1) Voice cloning or TTS tools for maintaining vocal consistency.
- 2) Experience with LoRA, ControlNet, or related fine-tuning techniques.
- 3) Background in advertising, post-production, or CGI.

Less than 30 hrs/week

Hourly

Less than 1 month

Duration

Expert

I am willing to pay higher rates for the most experienced freelancers

Contract-to-hire opportunity

This lets talent know that this job could become full time.

Learn more

Project Type: One-time project

You will be asked to answer the following questions when submitting a proposal:

Link to Portfolio or demos showing AI-based human models (image or video) where consistency across variations is evident.

Brief narrative of your process, tooling, and how you maintain consistency.

A note on how you'd collaborate to improve workflow and share knowledge.

Which generative AI tools have you used—like Stable Diffusion, ComfyUI, Veo-3, LoRA, or ControlNet? Describe your usage briefly

How do you ensure a character's face, voice, accent, and skin tone remain consistent across multiple outputs? Please share your approach (tools, automation, checkpoints)

Skills and Expertise

Mandatory skills

Video Commercial

Short Video Ad

AI-Generated Video

Audio Editing

Video Enhancement

Stable Diffusion

ComfyUI

Veo 3

AI Avatar Creation

Digital Human Modeling

Voice AI Agent Setup Specialist - RetellAI

Summary

We are looking for a skilled professional with extensive experience in setting up Voice AI agents, specifically using RetellAI. Your main responsibilities will include creating both inbound and outbound medical agents that effectively interact with patients and healthcare staff. The ideal candidate will have a strong understanding of voice technology in the

medical field and the ability to customize solutions for our needs. If you are passionate about improving healthcare communication through AI, we want to hear from you!

Priced: \$18 - \$40

Prompt Engineer (AI Animal Training Assistant)

Summary

 About the Project

We are building an AI-driven Animal Training Assistant – inspired by the methods of well-known animal trainers – that will work like ChatGPT:

Multilingual support (English, German, more languages will follow, etc.)

With memory to remember user progress and previous interactions

Motivating & supportive communication style

Personalized training plans automatically generated for each user's animal and situation

Our mission is to combine state-of-the-art prompt engineering with animal training expertise to create an intelligent companion for pet owners and professional trainers.

 Your Role

As a Prompt Engineer, you will:

Design and optimize prompt libraries for animal training scenarios (supportchat, dogs, cats, horses, cats, birds, etc.)

Create multi-turn conversations that feel natural, motivating, and supportive

Ensure the assistant can adapt to different languages & cultural contexts

Work on memory chains so the assistant remembers user history and adapts training plans over time

Collaborate with our AI team to fine-tune outputs into structured, individualized training programs

Develop evaluation frameworks (success metrics, safety guidelines, training accuracy)

Replit / Lovable App Building Workshop For Teens

Summary

Hi

We are looking for someone who is expert in Replit/Lovable to for 3 week series of workshops of app building.

We are looking for someone who can enjoy imparting knowledge to kids worldwide and align with our vision of giving a level playing field to all kids around the globe. We believe most kids are being shelled by their ecosystem, and we have to expose them to the world.

We are looking for someone who can enjoy the process, enjoy working with young brains and can make the workshops fun learning.

If this is something you can enjoy. Let's talk.

Best

Sam

AI Powered Full Stack Developer with AI Chatbot, AI Agent and AI Integration Experience

Summary

We are building an AI-powered SaaS platform and are looking for an experienced Full Stack Developer (individual or small team) to join us in developing the MVP and scaling it into a production-ready product.

The platform will leverage AI agents to automate workflows, handle user interactions, and provide intelligent recommendations. You should be comfortable working across the stack — from responsive UI to backend APIs, database design, and AI integration.

Responsibilities

Design and develop a scalable SaaS web application (frontend + backend).

Implement multi-tenant architecture with role-based access control.

Integrate AI agents (via OpenAI, Anthropic, LangChain, or similar frameworks).

Build workflows where AI agents assist users (e.g., handling queries, generating insights, or automating tasks).

Create dashboards, analytics, and reporting tools.

Ensure the application is secure, fast, and scalable.

Deploy on cloud platforms (AWS, GCP, or Azure).

Required Skills

Frontend: React / Next.js, Tailwind CSS (or similar modern UI frameworks).

Backend: Node.js (Nest.js / Express) or Python (FastAPI).

Database: PostgreSQL / MongoDB with strong relational design.

AI Integration: Experience with AI agents, LLM APIs (OpenAI/Anthropic), LangChain, vector databases (Pinecone, Weaviate, or pgvector).

Cloud & DevOps: AWS / GCP (serverless or containerized deployments with Docker/Kubernetes).

Strong understanding of secure SaaS development and API integrations.

Nice-to-Have

Experience building MVPs for SaaS platforms.

Background in workflow automation or chatbot/AI assistant development.

Knowledge of multi-language support, payments (Stripe), and subscription billing.

Deliverables

Functional SaaS MVP with AI agent integration.

Scalable database + API design.

Responsive and modern frontend UI.

Documentation for future development and maintenance.

We're excited to collaborate with someone who can combine solid full-stack engineering with AI-first thinking to bring this platform to life.

Priced: \$5,700.00

Upwork AI Developer Job Description

Summary We're building Project Z - a Solana trading bot + terminal that integrates directly into Discord & Telegram, surfacing on-chain + social signals (price, volume, holders, sentiment from X/Twitter) and letting users trade instantly. The MVP is already live (paper trading, Supabase + Python backend, Discord commands), but we need strong engineers to take it from a few users → thousands.

We're hiring two engineers:

1. **Backend Engineer (Trading Infrastructure)** • Optimize Supabase/Postgres schema and caching • Harden APIs for high-volume, low-latency trading • Extend demo/paper trading → live Solana trading (DEX APIs, wallets, transactions) • Add portfolio tracking, enforce limits (max positions, PnL sync, etc.) • Prepare backend APIs for future web terminal integration
2. **AI/LLM Engineer (Intel & Signals Layer)** • Improve prompt design and reduce generic outputs (grounding in cached tweets + on-chain data) • Explore RAG/vector DBs for better context grounding (Pinecone, Chroma, etc.) • Integrate additional data feeds (social/news) for sharper intel • Tune model usage for efficiency (OpenAI, Claude, Grok, or local fine-tuned models) • Own the LLM layer that powers intel, "More," and future "News" modules

Skills We're Looking For • Backend Engineer: Python, TypeScript, Supabase/Postgres, Solana/Web3.js, DEX API integration, low-latency trading bots, backend scaling • AI/LLM Engineer: Prompt engineering, RAG/vector DBs, LangChain/LlamaIndex, OpenAI/Claude APIs, Python, grounding outputs in structured data • Bonus (for both): Familiarity with Discord/TG bots, high-throughput systems

How We Work • Fast-paced, iterative build → test → improve. • Communication: Discord + async updates. • You'll work directly with the founder. • Looking for engineers who take ownership: not just code monkeys, but partners who think about performance, user experience, and scaling.

Why Join • Trading bots are one of the most profitable segments in crypto right now. •

Project Z is combining social trading + AI intel + instant execution with a flywheel (revenue

→ buyback). • You'll be building something users actually want, with direct line to market adoption and strong incentives to grow with us.

👉 To Apply: • Specify which role you're applying for (Backend or AI/LLM) • Share relevant past work (bots, trading infra, LLM pipelines, etc.) • Bonus: tell us how you'd make Project Z 100x sharper/faster

Priced: ~\$8800