Phase 1: Foundation Setup
- Google Workspace AI Activation: Enable Gemini features across Google Workspace applications
- Document Audit: Catalog existing Google Drive structure and document types
- Access Control Setup: Configure appropriate permissions for AI agent access
- Vector Database Selection: Implement Chroma for initial development with migration path to Pinecone for production scaling

Phase 2: Document Processing and Classification
- Automated Classification: Deploy Google Document AI Custom Document Classifier for the specified categories
- Content Extraction: Implement Gemini API for comprehensive document content analysis
- Metadata Enhancement: Apply automated tagging based on content analysis and organizational taxonomies
- Quality Assurance: Establish validation processes for classification accuracy

Phase 3: AI Agent Development
- RAG System Implementation: Build knowledge base using processed documents
- Query Interface Development: Create user-friendly interface for staff interactions
- Response Validation: Implement source citation and accuracy verification
- Training and Deployment: Provide staff training and gradual system rollout

Phase 4: Advanced Automation
- File Organization Automation: Implement Google Apps Script for automatic file reorganization
- Workflow Integration: Connect AI agent with grant writing and marketing workflows
- Historical Document Processing: Apply specialized OCR and heritage digitization techniques
- Continuous Improvement: Monitor performance and refine AI models based on usage patterns

Backend:
RAG : AI agent + fine-tune
Receive the files from the frontend

Frontend : linked to Drive

Query -> frontend -> backend (LLM: has access to Drive, fetch the right)

Core AI Platform

Google Workspace with Gemini Integration

- Native Google Drive integration
- Free Gemini features for nonprofits through Google for Nonprofits
- Highest accuracy for Google Drive document processing
- Enterprise-grade security and compliance

Document Extraction and Processing

Google Apps Script + Google Drive API
- Completely free with Google Workspace
- Direct programmatic access to all Drive files
- Supports batch processing and automation
- Native OCR capabilities for 200+ languages

AI Models for Content Analysis

Google Gemini 1.5 Flash (via API)
- Free tier available with generous limits
- Native PDF processing up to 1,000 pages
- Multimodal capabilities for text and images
- Best integration with Google ecosystem

Document Classification and Tagging

Google Document AI Custom Document Classifier
- Free tier with substantial monthly processing
- 96.5% accuracy for automated categorization
- Pre-trained models for common document types
- Minimal training data requirements

Automation Platform

Google Apps Script
- Completely free platform
- Native Google Workspace integration
- Cloud-based JavaScript environment
- Automated file organization by date/type/content

Vector Database for RAG System

Chroma
- Open-source and completely free
- Excellent LangChain integration
- Fast similarity search capabilities
- Python/JavaScript support

File Organization and Structure

AI File Pro (Open Source Alternative)
- Free Python-based solution
- Intelligent folder structure suggestions
- Metadata-based organization
- Customizable categorization rules

Workflow Integration

Google Workspace Flows
- Free with Google Workspace for eligible nonprofits
- Multi-step AI automation processes
- Custom AI agents ("Gems")
- Native integration across all Google services

Document Content Understanding

LayoutLM (Open Source)
- Free MIT license
- 94.4% accuracy for document classification
- Processes text, layout, and visual elements
- Commercial use permitted

Alternative Free Solutions by Category

For Basic File Organization
- **Google Apps Script** over paid alternatives like File Juggler
- **Native Google Drive search** over third-party indexing tools

For Document Processing
- **Google Document AI** over Azure Document Intelligence (free tier)
- **Gemini API** over OpenAI GPT models (cost efficiency)

For Automation
- **Google Workspace Flows** over Zapier (native integration)
- **Apps Script** over Microsoft Power Automate (no licensing costs)

For Classification
- **Google's Custom Document Classifier** over BERT implementations
- **Native Workspace AI** over third-party ML services

Implementation Priority

1. **Start with Google Apps Script** - Free foundation for all automation
2. **Enable Google Workspace AI features** - Immediate document analysis capabilities
3. **Implement Chroma vector database** - Free knowledge base foundation
4. **Deploy Google Document AI** - Automated classification within free tier
5. **Scale with Workspace Flows** - Advanced automation as needs grow