Roadmap: AI-Driven Cryptocurrency Due Diligence and Q&A System

1. Business Understanding

Objective:

- Build a due diligence platform for cryptocurrency and digital assets.
- Automate Q&A, risk assessment, and generate presentation slides.

Key Stakeholders:

- Cryptocurrency Investors
- Compliance Teams
- Crypto Enthusiasts

Desired Outcomes:

- Increase confidence in investments.
- Save time and effort through automation.
- Ensure scalability for document and query handling.

2. Data Science Framework

A. Data Understanding:

- Cryptocurrency project data: Whitepapers, ICO details, team information.
- Compliance and regulation documents.
- Market data: Prices, trends from APIs.
- Sentiment and social data: Twitter, Reddit, news articles.

B. Data Preparation:

- Parse and preprocess documents (PDFs, DOCX).
- Build a structured knowledge base.
- Collect labeled sentiment data.

C. Modeling:

- Q&A System: Train GPT-4 or similar on crypto documents.
- Risk Scoring Model: Combine sentiment analysis and risk factors.
- Document Understanding: Fine-tune models on whitepapers and compliance data.
- Presentation Generation: Use python-pptx for slide creation.

D. Deployment:

- Frontend: React or Streamlit for user interaction.
- Backend: FastAPI or Flask for integration.
- Cloud Services: AWS, Azure, or Google Cloud.

3. Roadmap for Building the System

Phase 1: Foundational Setup (1-2 weeks):

- Learn Python, APIs, and basics of AI modeling.
- Collect and preprocess whitepapers, compliance docs, and crypto data.

Phase 2: Build the Core Features (4 weeks):

- Document Understanding: Train models to read and extract information.
- Q&A System: Build chatbot for answering crypto-specific queries.
- Risk Analysis: Use sentiment analysis and scoring models.
- Presentation Generation: Automate slides with python-pptx.

Phase 3: Integration and Deployment (4 weeks):

- Create a user-friendly web interface.
- Set up backend and database.
- Deploy the application to the cloud.

4. Tools and Technologies

Programming and AI Libraries:

- Python: Pandas, NumPy, PyPDF2, python-pptx.
- Generative AI: OpenAI API, LangChain.

APIs and Data Sources:

- CoinGecko API, CoinMarketCap API for crypto data.
- Elliptic, Chainalysis for compliance.

Frontend Tools:

- Streamlit or React for interface development.

5. Example Timeline

Phase 1: Learn & Plan (1-2 weeks): Learn Python, APIs, and basic AI modeling.

Phase 2: Data Collection (2 weeks): Collect and preprocess data.

Phase 3: Core Features (4 weeks): Build document parser, Q&A, and risk models.

Phase 4: Presentation Generation (2 weeks): Automate PowerPoint creation.

Phase 5: Deployment (2 weeks): Create interface and integrate AI features.