AI Detection Using the Aparavi Data Toolchain

Overview of the Problem

Generative AI adoption is accelerating inside enterprises often without approval or visibility from security or compliance teams. Employees frequently paste sensitive information into public LLMs or generate business-critical content using AI without proper review.

My demo project, **AI Detection**, uses the **Aparavi Data Toolchain + AI stack** to automatically detect AI-generated content inside unstructured files across an enterprise's data lakes, file shares, or cloud repositories. The goal is to help organizations **gain visibility, assess risk, and enforce governance** on AI-created content.

How I Used the Data Toolchain

I integrated the Aparavi platform programmatically, creating a custom pipeline in Python that mimics how Aparavi automates classification workflows.

- Search for documents
- Retrieve document metadata
- Classify the input text using an AI model
- Write enriched metadata back to Aparavi

What I Built

My demo consists of a simple AI detector that works with:

- .pdf file input

The flow is simple:

1. User Inputs Text (PDF file)
2. LLM Evaluation (AI Fingerprinting)
3. Return JSON: { ai_generated, confidence, reason }

End-to-End Flow

1. Aparavi extracts text from the document
2. Aparavi sends the extracted text and file_id to the /classify endpoint
3. The /classify service runs the AI classifier
4. The service updates the file's metadata in Aparavi using the SDK
5. Aparavi displays AI detection results in the UI and dashboards

Why This Matters & Future Potential

AI detection is an emerging need in enterprise data governance. My approach demonstrates how Aparavi could evolve into an important **AI Governance Module**.

- Security teams need to detect unauthorized AI use.
- Legal teams must ensure contracts/policies are not auto-generated.
- Regulated industries require provenance tracking for AI content.
- Leaders need visibility into where employees rely on AI.

My demo could evolve into:

1. AI Provenance Tracking: Automatic labeling of documents created or edited by AI.
2. AI Safety Policies: Rules like flag documents with sensitive data generated by AI.
3. Integrate more formats: The project can integrate more formats like images, PDFs, and different file formats.

This shows how Aparavi can become a leader in **AI governance + unstructured data management**.

My Experience Using the Platform

My experience using the Aparavi Data Toolchain was surprisingly smooth, especially considering that I was integrating a custom LLM-based classifier with an enterprise-grade data governance platform. The product balances flexibility with structure in a way that makes building a full end-to-end Shadow AI Detection demo very achievable.

Conclusion

Aparavi exceeded my expectations in terms of **usability, scalability, and integration flexibility**. The platform enabled me to build a meaningful AI governance solution in a very short timeframe, without sacrificing architectural cleanliness or enterprise readiness.

What impressed me most is that Aparavi feels capable not just of organizing unstructured data, but of becoming a **core AI-risk intelligence platform** for modern enterprise.

It was a genuinely enjoyable platform to build on, and I came away with a strong sense of how the product could evolve and the value it can deliver.