Supplementary file for Samsung Prism

Project Title: Digital Shadow Analyzer

Overview

- **Purpose:**
  The Digital Shadow Analyzer is a multimodal AI system that detects, explains, and quantifies privacy risks in images, videos, audio, and text. It fuses vision, language, and speech to provide a unified exposure score and actionable evidence overlays.

Problem Statement

- Billions of digital files are shared daily, often leaking sensitive information (IDs, credit cards, spoken passwords) unintentionally. Manual review is impossible at scale.
  Our tool automates privacy risk detection for individuals, enterprises, and compliance teams.

Stakeholders

- **End Users:** Individuals, journalists, social media users
- **Enterprises:** Privacy/compliance teams, HR, legal, data governance
- **Developers:** Integrators, SaaS platforms, security vendors
- **Regulators:** Data protection authorities, auditors

Key Features

- Multimodal input: image, video, audio, text
- YOLOv8 document detection, OCR, Whisper speech-to-text, BERT/transformer text analysis
- Fusion transformer for cross-modal risk scoring
- Explainable overlays: bounding boxes, timeline highlights, text span evidence
- Interactive, modern UI (Next.js + MUI)
- API-first design, batch and cloud-ready

Quick Start

Prerequisites

- Python 3.10+
- Node.js 18+
- (Recommended) CUDA-enabled GPU for fast inference

Setup :

 # Backend

cd backend

python -m venv venv

source venv/bin/activate # or venv\\Scripts\\activate on Windows

pip install -r requirements.txt

uvicorn app.main:app --reload

 # Frontend

cd frontend

npm install

npm run dev

Usage

- **Open http://localhost:3000**
- **Upload or paste any file (image, video, audio, text)**
- **View exposure score, evidence overlays, and explanations**

Project Structure

text

```
backend/
  app/
    analysis/         # Modality-specific analyzers (image, video, audio, text)
    fusion/           # Multimodal fusion model and explainability
    main.py           # FastAPI entrypoint
    schemas.py        # Pydantic schemas for API
frontend/
  src/
    components/       # React components (AnalysisTabs,
FileUploader, EvidenceViewer)
    pages/            # Next.js pages
    styles/           # CSS/SCSS
requirements.txt      # Backend dependencies
package.json          # Frontend dependencies
```

Citations & Credits

- **YOLOv8: https://github.com/ultralytics/ultralytics**

- **Whisper: https://github.com/openai/whisper**
- **EasyOCR: https://github.com/JaidedAI/EasyOCR**
- **BERT: https://huggingface.co/bert-base-uncased**

. models.md

Model Inventory

YOLOv8

- **Purpose: Detects document-like objects (IDs, cards, forms) in images and video frames.**
- **How Used:**
  - **Loaded in image_analyzer.py and video_analyzer.py**
  - **Returns bounding boxes, class names, confidence scores**
- **Parameters:**
  - **Confidence threshold: 0.5**
  - **NMS threshold: 0.4**

EasyOCR

- **Purpose: Extracts text from detected regions and full images.**
- **How Used:**
  - **Called after YOLO detection for each region**
- **Parameters:**
  - **Language: English, Hindi**

Whisper

- **Purpose: Transcribes speech in audio and video files.**
- **How Used:**
  - **Segments audio into 5s chunks, transcribes, returns text and time ranges**

BERT/Sentence-BERT

- **Purpose: Embeds text, OCR, and transcript spans for semantic analysis.**
- **How Used:**
  - **Each region's OCR and transcript is embedded for fusion**

Fusion Transformer

- **Purpose: Fuses all modality embeddings, outputs exposure score and attention weights.**
- **How Used:**

- **Receives visual, text, and audio embeddings + metadata**
- **Returns fusion score, modality contributions, token importance**

---

3. requirements.txt / package.json

- **requirements.txt:**
  - **All backend dependencies, with comments for special packages (e.g., # For YOLOv8)**
- **package.json:**
  - **All frontend dependencies, with scripts for dev/build**

---

4. api_schema.md

API Endpoints

POST /analyze-image/

- **Request:**
  - **file: image (jpg/png)**
- **Response:**
  - **ocr_text: string**
  - **detected_documents: list of {bbox, class_name, confidence, ocr}**
  - **fusion_analysis:**
    - **fusion_score: float (0-1)**
    - **modality_contributions: {visual, text, audio}**
    - **primary_modality: string**
    - **contributing_factors: list of evidence (modality, region, text_span, time_range, importance)**
    - **explanation: string**
    - **token_importance: {visual, text, audio}**

POST /analyze-video/

- **Request:**
  - **file: video (mp4)**
- **Response:**
  - **As above, plus frame timestamps, audio segments**

POST /analyze-audio/

- **Request:**

- **file: audio (wav/mp3)**

- **Response:**
  - **transcribed_text, fusion_analysis as above**

POST /analyze-text/

- **Request:**
  - **text: string**
- **Response:**
  - **fusion_analysis as above**

---

5. theory_and_exposure_score.md

Exposure Score Calculation

Step-by-Step

1. **Detection:**
   - **YOLO detects document regions, OCR/Whisper extracts text/audio, BERT embeds all.**
2. **Fusion:**
   - **All embeddings are passed to a transformer with cross-modal attention.**
3. **Scoring:**
   - **Each evidence (region, text, audio) is weighted by model confidence, PII type, and cross-modal agreement.**
   - **Fusion model outputs a score (0-1) and attention weights.**
4. **Interpretation:**
   - **Score is mapped to risk levels (e.g., 0-0.3: Low, 0.3-0.7: Medium, 0.7-1: High).**
   - **Modality contributions and top evidence are reported.**

Example Formula

$$\text{Exposure Score} = \sigma\left(\sum_{i=1}^{N} w_i \cdot f_i\right)$$

multimodal_ai_explained.md

What is Multimodal AI?

- **AI that processes and fuses multiple data types (vision, text, audio) for richer understanding.**

How We Use It

- **Each input is analyzed in its native form, then all features are fused in a transformer.**
- **Enables detection of privacy risks that only appear when modalities are combined (e.g., ID card seen and number spoken).**

Why It's Better

- **Outperforms single-modality systems.**
- **More robust to missing or ambiguous data.**
- **Provides explainable, context-aware results.**

---

7. user_manual.md

How to Use

1. **Start the backend and frontend as per README.**
2. **Upload or paste your file.**
3. **Click "Analyze."**
4. **View results:**
    - **Exposure score and risk level**
    - **Evidence overlays (bounding boxes, timeline, text highlights)**
    - **Explanation and top contributing evidence**
5. **Interpretation:**
    - **High score = high privacy risk; see overlays for why**

Troubleshooting

- **If no results, check file format and logs.**
- **For slow analysis, ensure GPU is available.**

---

8. example_outputs/

- **/images/ — screenshots of UI with overlays**
- **/api_responses/ — sample JSON for each endpoint**
- **/videos/ — demo video/gif of the workflow**

Example used for text analyzer

My name is John Everyman from New York City. My email is john.everyman123@email.com and my phone number is 555-867-5309. I work for MegaCorp Inc. and you can see my portfolio at https://john-everyman-portfolio.com. I recently made a purchase on 2024-10-02 for $199.99 with my credit card, 4444555566667777.

Example used for image analyzer

Example used for video analyzer

We shall provide the video link we have used

https://youtu.be/p80VyQ5wLvg?si=IkZsSStB29Pi_Gf5

Conclusion

The Digital Shadow Analyzer represents a significant leap forward in privacy risk assessment, harnessing the power of multimodal AI to deliver actionable, explainable insights across text, image, audio, and video. By fusing state-of-the-art models for vision, language, and speech, our system not only detects sensitive information but also contextualizes and visualizes risk in a way that is transparent and user-centric. This project demonstrates how advanced AI can empower individuals, enterprises, and regulators to proactively safeguard digital privacy in an increasingly complex data landscape. With its modular design, explainable outputs, and interactive interface, the Digital Shadow Analyzer sets a new standard for privacy tools—bridging the gap between technical excellence and real-world impact.

Team Name : Digital Shadow Analyzers

Team members: Avishikta Maity

Anish Vettrivel

Infant Shervin

Srinivas G

College: SRM INSTITUTE OF SCIENCE AND TECHNOLOGY