🔍 **Problem 2: Poor Data Quality & Inconsistent Formats**

🧩 **Problem Definition**

The police face severe challenges in identifying missing and unidentified persons due to **low-quality images, incomplete data entries, and inconsistent database formats** across various agencies such as hospitals, NGOs, shelters, and mortuaries.
The lack of standardized data reduces the efficiency of AI recognition models and slows down the process of cross-verification and identification.

⚙️ **Detailed Analysis**

1. Nature of the Problem

- Each department maintains its own data structure — Excel sheets, PDFs, scanned documents, handwritten notes, or digital photos stored in non-standardized ways.

- Photos captured under poor lighting or degraded image conditions (especially of bodies) further reduce facial recognition accuracy.

- Metadata such as *time, location, and context* are often missing or inconsistently recorded.

- The absence of a unified schema causes failure during data merging and model training, making automation difficult.

💡 **Proposed AI & Data Engineering Solutions**

Challenge	Solution	Tech Stack / Tools
Blurry / unclear images	AI-powered image enhancement using super-resolution models.	OpenCV, ESRGAN, TensorFlow / PyTorch
Different data formats (CSV, Excel, PDF, text)	Automated data cleaning and format conversion.	Python (Pandas, NumPy), Apache NiFi / Airflow for ETL
Missing metadata (date, place, etc.)	AI-based metadata extraction and labeling.	spaCy NLP, EXIF data readers

Challenge	Solution	Tech Stack / Tools
Lack of unified schema	Creation of a standardized JSON schema for all participating departments.	PostgreSQL / MongoDB, JSON Schema
Manual errors and duplication	Auto-validation pipeline to check and flag incomplete or duplicate records.	FastAPI middleware + Python validation scripts

🔬 **Feasibility–Viability–Desirability Analysis**

✅ **Feasibility (⭐ ⭐ ⭐ ⭐ ☆ – 4/5)**

Reasoning:

- Technically feasible with readily available open-source tools and machine learning frameworks.

- Data pipelines can be built using **Python, FastAPI, and cloud-based ETL tools** without requiring proprietary systems.

- Image enhancement and data validation models are well-documented and implementable.

- Integration may require moderate coordination among departments but no major structural changes.

Conclusion:
Technically and operationally **feasible** with scalable architecture and minimal dependency on expensive technologies.

💰 **Viability (⭐ ⭐ ⭐ ⭐ ☆ – 4/5)**

Reasoning:

- Uses **open-source frameworks**, reducing software licensing costs.

- Cloud deployment (AWS GovCloud / Azure Public Sector) offers **pay-as-you-go** scalability.

- Automated cleaning and validation pipelines significantly **cut manual labour and time costs**.

- Fits within the **Digital India and CCTNS modernization** budget structures.

- Maintenance involves minimal recurring costs once the initial system is deployed.

Conclusion:
Economically **viable** and sustainable for long-term use by state or national authorities.

❤️ Desirability (⭐ ⭐ ⭐ ⭐ ⭐ – 5/5)

Reasoning:

- **Police** gain reliable, standardized, and searchable data that speeds up investigations.
- **Hospitals, NGOs, and shelters** benefit from easier collaboration and quick reporting.
- **Families** receive faster identification results and emotional closure.
- Enhances **public trust** and demonstrates social responsibility through the ethical use of AI.

Conclusion:
Socially and emotionally **highly desirable** as it directly supports human rights, justice, and empathy-driven policing.

🌐 Overall Summary

Category	Key Takeaways
Feasibility	Fully achievable with modern AI and ETL tools; scalable and integrable with current databases.
Viability	Cost-effective, open-source, and aligned with ongoing government digital initiatives.
Desirability	Deeply human-centered, promoting trust, speed, and social welfare.
Tech Stack Summary	Python, Pandas, NumPy, OpenCV, ESRGAN, TensorFlow / PyTorch, FastAPI, PostgreSQL / MongoDB, Apache Airflow / NiFi, spaCy.
Overall Rating	⭐ ⭐ ⭐ ⭐ ☆ *(4.3 / 5)*

🧠 Final Insight

The AI-based data standardization and enhancement pipeline transforms scattered, unreliable data into **high-quality, unified, and searchable information**, empowering the police and connected institutions to identify missing and unidentified persons **faster, more accurately, and more ethically**.