

# Milestone 1 — AI Video Collector (Data Structuring & Enrichment Foundation)

## Objective

Transform the existing WhatsApp export (~1,200 links) into a **fully structured, enriched dataset** containing every AI- or tech-related video, article, or post — normalized, categorized, and ready for automation.

## Core Goals

### 1. Parse & Structure

- Extract all links and metadata from the WhatsApp export.
- Standardize fields:  
`Date | Name of Channel | Category | Title | URL link | Description | Platform | Tags | Followers | Platform |`
- Detect platform type automatically (YouTube, TikTok, Instagram, X, websites, etc.).
- Categorize content into predefined categories: *AI Creative/How-To, AI Agents, AI Tools, AI News & Updates, Business Strategy, Tech/Startups, Other.*

### 2. Data Quality Audit & Cleanup

- Identify missing data per platform (e.g., missing titles, channel names, or follower counts).
- Detect YouTube IDs, top domains, and potential duplicates.
- Produce a clear audit report for enrichment priorities.

### 3. Automated Enrichment via APIs

- **YouTube API v3 (Google Cloud):**
  - Pull Channel Name, Channel URL, Subscribers, Video Title (official), Duration, Publish Date.
- **TikTok oEmbed & Instagram Graph oEmbed:**
  - Retrieve creator name and video/post title (followers optional).
- **Web/News Open Graph scrape:**
  - Fetch meta titles for websites or articles.

### 4. Output / Integration

- Deliver a single, clean CSV (Google Sheets-compatible) sorted newest → oldest.
- Each entry includes platform, category, and enrichment placeholders for next-phase AI summaries.

## Deliverables

- `WHATSAPP_AI_Video_Library_2025-10-12_v2_enriched.csv`
- Audit report (Missing fields + Platform coverage)
- YouTube and Social Enrichment Scripts ready to run

## Estimated Dev Time (Setup & Testing)

---

## Milestone 2 — AI Video Intelligence (Summarization & Automation)

### Objective

Add **AI-generated summaries, insights, and full automation** on top of the structured dataset to create a living “AI Video Library.”

### Core Goals

#### 1. Content Summarization & Classification

- Auto-generate:
  - 5-sentence summary per video/post
  - 4–8 bullet key points
- Enrich categories dynamically (e.g., *AI News, Agent Tech, Tools & Apps, Research, Ethics/Safety, etc.*).
- Add relevant tags and keywords (models, tools, companies).

#### 2. Automation Pipeline

- Integrate with **Google Sheets API** or **Notion / Airtable** for live updates.
- Optional: add **Zapier / Make** or **Next.js API** endpoints for on-demand enrichment.

- Allow WhatsApp “Share-to-Me” automation — new links auto-append to sheet, enriched + classified.

### 3. Intelligence Enhancements

- Add “Top Creators,” “Trending Topics,” and “Recent AI Releases” views.
- Include follower-based ranking or recency-based scoring.
- Generate weekly or monthly digests automatically.

### Deliverables

- `AI_Video_Library_FINAL.csv` (or connected Google Sheet)
- Summaries + Key Points + Tags columns populated
- LLM-based enrichment script (batch or API)
- Optional: dashboard or API endpoints for live enrichment