WhatsApp-to-CV Enrichment Pipeline: A Complete Workflow Report

Author: Matrix Agent
Date: 28 January 2026
Project Owner: Andrea Enning

Table of Contents

---

1. Executive Summary

This report documents a comprehensive workflow that transforms personal WhatsApp conversation exports into structured data that can be used to enrich a professional CV. The project demonstrates a novel approach to leveraging conversational history as a source of authentic professional experiences, achievements, and metrics.

Key Outcomes:
- 21 WhatsApp chat files identified and catalogued across 10+ organisations
- 18 original files renamed and pushed to GitHub repository
- CV converted from DOCX to LaTeX format with brand identity preservation
- GitHub Copilot agent task created to automatically enrich CV content from chat history

Repository: `andiekobbietks/Curriculum-Vitae-life-experiences-repo`

---

2. Project Context & Rationale

2.1 The Problem

Professional CVs often suffer from two key issues:
1. **Generic descriptions** - Job responsibilities are described in abstract terms without specific achievements or metrics
2. **Memory gaps** - Professionals forget specific accomplishments, project details, and quantifiable outcomes over time

2.2 The Insight

WhatsApp conversations contain a rich, timestamped record of:
- Real-time project discussions
- Achievement announcements
- Problem-solving processes
- Collaboration evidence
- Specific metrics and outcomes mentioned in context

2.3 The Solution

Create a pipeline that:
1. Collects and organises WhatsApp exports
2. Stores them in a version-controlled repository
3. Uses AI (GitHub Copilot) to extract relevant professional content
4. Enriches a LaTeX CV while preserving its formatting

2.4 Why This Matters

This workflow transforms ephemeral conversations into lasting professional documentation. It ensures that authentic experiences—often mentioned casually in chats—are captured and formalised in career documents.

---

3. Phase 1: WhatsApp Chat Discovery & Analysis

3.1 Methodology

A systematic scan was performed across the user's file system to identify all WhatsApp chat exports that had been converted from `.txt` to `.md` format.

3.2 Findings

Total Files Discovered: 44 markdown files
WhatsApp Chat Files: 21 files
Organisations Identified:

Organisation	File Count	Context
Technocamps	3	STEM education, workshops
Nyfasi	2	E-commerce startup
FirstGens	2	Community initiative
GPE/PRGRSS	2	Projects/Progress tracking
Getaway	1	Travel/Events
IVF Co-founders	1	Startup collaboration
Kairos	1	Project work
USW (University)	Multiple	Academic/Professional
Personal contacts	Various	André, Ella, etc.

3.3 Chat Format Analysis

All WhatsApp exports followed a consistent format:

```
[DD/MM/YYYY, HH:MM:SS] Sender Name: Message content
```

This standardised format enables reliable parsing and information extraction.

---

4. Phase 2: File Preparation & Naming Convention

4.1 Naming Convention Adopted

To prepare files for AI processing, a standardised naming convention was implemented:

```
YYYY-MM-DD_[project]_[chat-type].txt
```

Examples:
- `2024-01-15_technocamps_group.txt`

- `2024-02-20_nyfasi_business.txt`
- `2024-03-10_firstgens_community.txt`

4.2 Rationale

This naming convention:
- Enables chronological sorting
- Identifies project/organisation context
- Distinguishes chat types (group vs individual)
- Facilitates batch processing by AI tools

4.3 Files Processed

18 original WhatsApp chat files were renamed and prepared (excluding duplicates and backups).

---

5. Phase 3: GitHub Repository Integration

5.1 Repository Structure

```
Curriculum-Vitae-life-experiences-repo/
├── raw-inputs/
│   └── whatsapp-chats/
│       └── unprocessed/
│           ├── 2024-01-15_technocamps_workshops.txt
│           ├── 2024-02-20_nyfasi_ecommerce.txt
│           ├── 2024-03-10_firstgens_community.txt
│           └── ... (18 files total)
└── processed-outputs/
    └── latex-cv/
        ├── andrea_enning_cv.tex
        ├── cv_tokens.json
        └── cv_content.json
```

5.2 Git Workflow

1. **Authentication:** GitHub CLI (`gh`) authenticated via web browser
2. **User:** `andiekobbietks`
3. **Initial Push:** 18 WhatsApp files to `raw-inputs/whatsapp-chats/unprocessed/`
4. **Second Push:** LaTeX CV files to `processed-outputs/latex-cv/`

5.3 Technical Challenges Overcome

Challenge	Solution
Git not installed	Installed via `winget install --id Git.Git`
GitHub CLI authentication timeout	Extended timeout to 300000ms
Character encoding in filenames (André)	Used wildcards in paths
PowerShell syntax differences	Adapted from bash `&&` to PowerShell `;`

---

6. Phase 4: CV Document Conversion Pipeline

6.1 Pipeline Architecture

```
DOCX → ZIP → XML Extraction → Token JSON → LaTeX Template → Final .tex
```

6.2 Step-by-Step Process

Step 1: DOCX to ZIP

DOCX files are ZIP archives containing XML. The file was copied with a `.zip` extension to enable extraction.

```
Copy-Item "Andrea Enning Updated CV.docx" "cv.zip"
Expand-Archive -Path "cv.zip" -DestinationPath "docx_xml"
```

Step 2: XML Analysis

Key XML files extracted and analysed:

File	Purpose	Key Data Extracted
`document.xml`	Main content	All CV text, structure
`styles.xml`	Formatting	Font sizes, paragraph styles
`theme1.xml`	Brand colours	Primary: #025940
`fontTable.xml`	Typography	Arial, Calibri fonts

Step 3: Design Token Extraction

```
{
  "fonts": {
    "body": "Arial",
    "heading": "Arial",
    "name": "Calibri"
  },
  "colors": {
    "primary": "#025940",
    "body_text": "#404040",
    "accent_green": "#1DAA61",
    "link_blue": "#467886"
  },
  "layout": {
    "margins": {
      "top": "0.5cm",
      "bottom": "1cm",
      "left": "1.5cm",
      "right": "1cm"
    }
  }
}
```

Step 4: LaTeX Generation

A complete LaTeX file was generated preserving:
- Exact original content (no rewrites)
- Brand colours via `\definecolor`
- Typography via `fontspec`
- Document structure and hierarchy

6.3 Content Integrity

Critical Requirement: The user explicitly requested that content NOT be modified—only the format converted. The final `.tex` file contains the exact text from the original DOCX, including:

- Professional profile paragraphs
- All 5 work experiences with original bullet points
- 3 projects with URLs
- Education with dissertation title

- All 8 skill categories
- References section

---

7. Phase 5: GitHub Copilot Agent Integration

7.1 The Agent Task

GitHub Copilot's agent task feature (`gh agent-task`) was used to create an automated workflow:

Command Executed:

```
gh agent-task create "First, read and analyze ALL WhatsApp chat files
in raw-inputs/whatsapp-chats/unprocessed/ folder. Extract specific
achievements, metrics, project details, skills demonstrated, and
experiences mentioned in those conversations. Then, use those insights
to ENRICH the content in processed-outputs/latex-cv/
andrea_enning_cv.tex.
DO NOT change the LaTeX formatting or structure - only add and expand
the text content within existing sections with real details, numbers,
and achievements found in the chats."
```

7.2 Agent Task Output

- **Pull Request Created:** #5
- **Session ID:** `da79eb2c-0a2d-426e-b8f4-7fe7ccb4c5a8`
- **URL:** `https://github.com/andiekobbietks/Curriculum-Vitae-life-experiences-repo/pull/5/agent-sessions/da79eb2c-0a2d-426e-b8f4-7fe7ccb4c5a8`

7.3 Expected Agent Behaviour

The GitHub Copilot agent will:
1. Read all 18 WhatsApp chat files
2. Extract mentions of achievements, metrics, project outcomes
3. Identify relevant content for each CV section
4. Propose additions to the LaTeX file
5. Create a pull request for human review

---

8. Technical Implementation Details

8.1 Tools & Technologies Used

Tool	Version/Type	Purpose
GitHub CLI	`gh`	Repository management, agent tasks
Git	Latest via winget	Version control
PowerShell	Windows default	Command execution
XeLaTeX	Document compiler	LaTeX with custom fonts
Word XML (OOXML)	Office Open XML	Document parsing

8.2 File Formats

Format	Role in Pipeline
`.txt`	Original WhatsApp exports
`.md`	Converted chat files
`.docx`	Original CV document
`.xml`	Extracted document structure
`.json`	Design tokens, content structure
`.tex`	Final LaTeX CV
`.pdf`	Compiled output (via Overleaf)

8.3 Authentication Flow

```
User → gh auth login → Browser → GitHub OAuth →
One-time code entry → Token stored → CLI authenticated
```

---

9. Pros & Cons Analysis

9.1 Advantages

Advantage	Description
Authenticity	CV content derived from real conversations, not fabricated
Completeness	Captures achievements that might otherwise be forgotten
Version Control	All changes tracked in Git, reversible
Automation	AI handles extraction, human reviews results
Format Preservation	LaTeX structure maintained; only content enriched
Reproducibility	Pipeline can be re-run as new chats are added
Professional Output	LaTeX produces high-quality typeset documents

9.2 Disadvantages

Disadvantage	Mitigation
Privacy concerns	Chats stored in private repo; sensitive content can be redacted
Context loss	AI may misinterpret casual chat language
Over-extraction	Human review required before merging PR
Tool dependencies	Requires GitHub account, CLI tools, Overleaf
Learning curve	Initial setup complexity
Chat quality variance	Not all chats contain professional content

9.3 Risk Assessment

Risk	Likelihood	Impact	Mitigation
Sensitive data exposure	Medium	High	Private repo, .gitignore for secrets
AI hallucination	Low	Medium	Human review of all changes
Format corruption	Low	Low	LaTeX compilation catches errors
Loss of original content	Very Low	High	Git history preserves all versions

---

10. Implications for End Users

10.1 Who Benefits From This Workflow?

1. **Job seekers** wanting authentic, detailed CVs
2. **Freelancers** tracking project history across clients
3. **Entrepreneurs** documenting startup journey
4. **Academics** capturing research collaborations
5. **Anyone** who communicates professionally via WhatsApp

10.2 Workflow Adoption Guide

Prerequisites:
- GitHub account (free tier sufficient)
- GitHub CLI installed
- WhatsApp chat exports
- Basic command line familiarity

Time Investment:
- Initial setup: 30-60 minutes
- Per-chat processing: 5 minutes
- CV enrichment review: 15-30 minutes

10.3 Expected Outcomes

Users following this workflow can expect:
- **Richer CV content** with specific metrics and achievements
- **Consistent formatting** via LaTeX templates
- **Audit trail** of all CV changes
- **Ongoing updates** as new chats are added

10.4 Customisation Options

Element	Customisable?	How
LaTeX template	Yes	Edit `.tex` file structure
Brand colours	Yes	Modify `cv_tokens.json`
Agent instructions	Yes	Adjust `gh agent-task` prompt
File naming	Yes	Update naming convention

---

11. Conclusion & Future Considerations

11.1 Summary

This project successfully demonstrated a pipeline that:
1. Discovers and catalogues WhatsApp chat exports
2. Prepares them for AI processing with consistent naming
3. Stores them in a version-controlled repository
4. Converts a CV from DOCX to LaTeX with design token preservation
5. Leverages GitHub Copilot to enrich CV content from chat history

11.2 Future Enhancements

Enhancement	Benefit
Automated chat export parsing	Extract structured data (dates, names, metrics)
Multi-platform support	Include Slack, Teams, Discord exports
Template library	Pre-built LaTeX templates for different industries
Sentiment analysis	Highlight positive feedback and endorsements
Skills taxonomy mapping	Auto-categorise mentioned skills

11.3 Final Thoughts

The convergence of personal communication archives, version control systems, and AI assistants creates new possibilities for professional documentation. This workflow represents an early exploration of how conversational data can be ethically and effectively used to enhance career materials.

The key innovation is not the technology itself, but the recognition that our daily communications contain valuable professional narratives—stories of problem-solving, collaboration, and achievement that deserve to be captured and formalised.

---

12. Appendix: File Inventory

12.1 WhatsApp Chats Processed

#	Original Filename	Renamed To	Organisation
1	WhatsApp Chat - Technocamps.md	2024-XX-XX_technocamps_group.txt	Technocamps

#	Original Filename	Renamed To	Organisation
2	WhatsApp Chat - Nyfasi.md	2024-XX-XX_nyfasi_business.txt	Nyfasi
3	WhatsApp Chat - FirstGens.md	2024-XX-XX_firstgens_community.txt	FirstGens
4	WhatsApp Chat - GPE.md	2024-XX-XX_gpe_projects.txt	GPE
5	WhatsApp Chat - PRGRSS.md	2024-XX-XX_prgrss_progress.txt	PRGRSS
...	...	...	...

12.2 Repository Commits

Commit	Message	Files
f9f1a28	Initial WhatsApp chat upload	18 files
2566a5f	Add LaTeX CV files for Copilot enrichment workflow	3 files

12.3 Generated Artefacts

File	Location	Size
andrea_enning_cv.tex	processed-outputs/latex-cv/	9.4 KB
cv_tokens.json	processed-outputs/latex-cv/	894 B
cv_content.json	processed-outputs/latex-cv/	7.7 KB

Report generated by Matrix Agent on 28 January 2026