

A Persona Proxy System Overview

Core Concept

A-Proxy has two archival goals:

1. **Persona-Driven Web Crawling:** Develop detailed persona representations through cumulative chat conversations to enable targeted web crawling from specific perspectives
2. **LLM Evolution Documentation:** Archive the chat interactions themselves to study how LLMs develop over time and respond to increasing personalization

The system uses chat conversations to progressively develop personas across four data categories.

The Feedback Loop

1. Persona Generation

- o Archivists create initial personas across four categories (with LLM assistance):
 - **Demographic:** location, language, age, gender, education, income, occupation
 - **Psychographic:** interests, values, attitudes, lifestyle, personality, opinions
 - **Behavioral:** browsing habits, purchase history, device usage, social media activity
 - **Contextual:** time of day, device type, browser, connection speed. These can be inferred by the LLM depending on the type of journey (shopping, research, marketing, the persona is engaged in, captured as journeys and waypoints.)

2. Persona Development through Chat

- o Archivists chat with personas to develop their attributes over time or give directives in a certain format (you have a job in a restaurant on 36th Street).
- o The relevant features are reified (attempting to build the components browsers and content delivery systems use to target user attributes)
- o Conversations progressively build out the four data categories

- Chat content informs crawling targets and browsing patterns
- Each conversation refines the persona's perspective for more targeted archiving

3. Targeted Crawling (Waypoints)

- Archivists crawl websites from persona perspectives. Personas are directed (through prompt refinement) to focus their conversations on the websites that interest them).
- Browser variables reflect persona characteristics (location, language, user-agent), cookies set context, and synthetic history files represent past activity
- Crawling sessions capture how different users might experience the web

4. Dual Archival Value

- **For Web Archiving:** Refined personas enable crawls from specific perspectives
- **For LLM Research:** Preserved chats document:
 - How LLMs respond to increasing personalization
 - Changes in model behavior as versions improve
 - Benchmarking of responses based on persona characteristics

Innovation

The system uniquely combines:

- **LLM Memory Techniques:** Preserving conversational context to build persona depth
- **Behavioral Modeling:** Translating personality into actual browsing patterns
- **Archival Value:** Each persona becomes a preserved representation of user behavior patterns

Use Cases

- Targeted web archiving from diverse perspectives
- Documenting how different demographics might experience websites
- Preserving persona-LLM interactions for future research

- Studying LLM behavioral changes over time
- Creating archival records that reflect multiple viewpoints