By Philip John / @Sexadegen

ragmilabs.ai

Executive Summary

This paper outlines a Human RAG (Retrieval Augmented Generation) solution that transforms human data into an active retrieval layer, allowing AI agents to better understand a human source without massive prompt overhead.

A major step forward in human understanding allows the AI agents to shift from just being an assistant to a companion, opening the door to a range of previously unachievable scenarios including:

- Deep understanding of the agent's human controller
- Long-term memory grounded in real human experience
- The use of AI agents to deliver episodic memories to dementia sufferers
- A human avatar for the agent, maybe a deceased relative or a fictional companion
- Emotional continuity across long-running interactions

This migration—from AI Assistant to AI Companion—represents a foundational shift in human-AI interaction and presents the opportunity to move a step further by using the Human RAG to give the agent its own identity and memories.

Introduction: The Need for Human Context

Since the launch of OpenClaw there has been an explosion of people creating their own personal AI assistants. Every day there are more applications being created, but a common observation is that the relationship between the user and the agent is very much master and slave.

For many people, the concept of the perfect AI assistant is JARVIS, the fictional character from the Iron Man movies. JARVIS does what it's asked but has a relationship with Tony Stark that is as much companion as assistant. JARVIS ***knows*** Tony, so the question is how do we enable our AI agents to know their users in the same way?

The choice is simple, we either get the agent to ask all the right questions and remember the answers, or the user teaches the agent in a way that the information is always available.

Getting the agent to ask the right questions has some merit, but it is brittle and exhausting. It is difficult for the agent to know which questions to ask, which are more important than others, and it is a tedious process for the user to be constantly answering questions.

The obvious solution is for the user to teach their agent all about themselves. LLMs use Retrieval Augmented Generation (RAG) to provide specialist knowledge on a particular subject. In the context of an AI agent, Human RAG provides the specialist knowledge of the person controlling the agent, or taking this a step further, it is possible to use Human RAG to give an agent its own set of memories to shape its personality and interests in a particular direction.

What the agent needs to know depends greatly on what you intend to use it for. If you just want your OpenClaw assistant to know that you prefer pizza to pasta or the date of your partner's birthday, a simple markdown list of preferences and facts will suffice. If you want your AI agent to be as close to JARVIS as you can get, then you need to tell it your life story. It needs to know who you are, where you've been, who you know, what you've done and what you like.

It's a significant task. You've probably been to a lot of places, done many things, and with many people. You may have shared interests that you want your AI agent to become expert in, and you may want your agent to have access to the relevant digital media for all these things.

The good news is that you can start small and build incrementally, and as you add new experiences, you can keep your agent up to date.

RAG**MI** allows you integrate all this information into a single application and then handles the export into formats that are digestible by AI agents.

RAG**MI** for Mac

Creating the user information that can be used by an AI agent is a massive mind dump. It needs an application that can pull together multiple different kinds of digital media and allows the user to create an expandable life story at the level of detail that they want.

RAG**MI** for Mac manages this process. The application allows you to compile all the information for your AI agent in a way that is enjoyable to use. It allows you to import audio, video, photos, documents and social media and create coherent timelines around them.

You can create all your personal and business connections and weight them accordingly so that your agent knows who's important and who's less so.

Once you've got sufficient information you can choose your export options and create files that can be easily consumed by OpenClaw or LLMs.

Folio Structure

Each project within the RAG**MI** application is known as a ***folio***, a container for all the media, stories, connections and events related to that subject. Current AI systems struggle with personal data because they lack a significance filter. To an LLM, a receipt for a coffee and a receipt for a wedding ring are just two financial transactions. RAG**MI** solves this by introducing a processing pipeline that embeds human weighting directly into the data structure before the AI ever sees it.

Digital Media

Some memories are simply defined by what we can remember, but in most cases, there is some form of media to anchor a story. RAG**MI** imports a wide range of digital media—photos, videos, audio, documents and social media archives, including recreation of WhatsApp chats, which can often perfectly encapsulate an event.

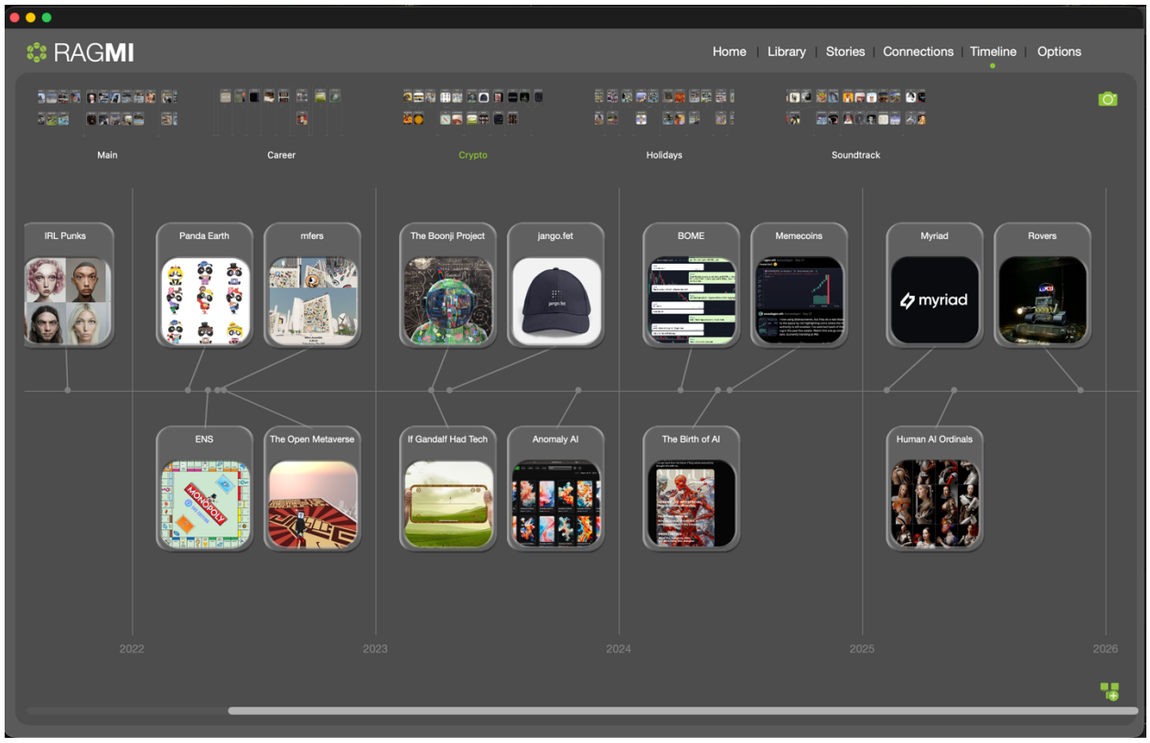
Every data point can be pinned to a specific moment or duration on one of multiple timelines.

Stories, Connections and Events

These are the wrappers that can be applied to any mixture of digital media to provide context to something or someone. Users can assign significance weighting to specific stories, events and people, ensuring that when an AI agent has a limited context window, it prioritizes high impact memories over trivial facts, reducing the noise-to-signal ratio.

Timelines

RAG**MI** presents stories and events chronologically in the form of multiple timelines. A single timeline for a life story can get cluttered very quickly so RAG**MI** supports the option to have a main timeline and multiple others for specific purposes, e.g. career-related, travel, special interests etc.



Human Context Protocol

To solve the bottleneck of agent context, RAG**MI** introduces the Human Context Protocol (HCP).

Unlike traditional data exports which simply dump raw files, HCP is a dual-layer standard designed to bridge the gap between structured human history and active AI cognition. It ensures that personal data is not only portable and durable but immediately intelligible to autonomous agents.

The protocol consists of two synchronized layers:

The Factual Layer (JSON-LD)

The foundation of the protocol is factual data. Built on JSON-LD, this layer preserves the strict hierarchy, relationships, and weights of the user's life data.

- Universal Portability: Ensures data can move between different AI agents or platforms without loss of fidelity.
- Long-Term Survivability: Decouples the user's history from any single vendor, allowing future agent architectures to consume legacy context without re-authoring.
- Graph Integrity: Maintains the raw connections between people, events, and timelines, ensuring the "who, what, and when" remain machine-readable.

The Narrative Layer (Semantic Markdown)

To enable immediate, high-fidelity interaction, HCP includes a standardized Narrative Layer. This layer transforms structured data into natural language, optimized for RAG ingestion.

This is based on narrative serialization, metadata injections and weighting.

By unifying the factual and narrative layers, the Human Context Protocol creates a living digital legacy: a permanent archive for the user, and a preference-aware cognitive map for immediate, significance aware execution by AI agents.

Folio Security

Concentrating a human life into a single dataset creates an unprecedented security target. Unlike standard corporate data, a breached Human RAG can reveal psychological, relational, and emotional vulnerabilities.

To mitigate this, RAG**MI** adopts a local-first architecture that ensures the user retains absolute sovereignty over their data. Privacy begins with where the computation happens. RAG**MI** operates entirely on the user's local machine (Mac), not in the cloud.

The conversion of raw media into the HCP occurs locally, and the resulting JSON-LD is stored as a local file, not on RAG**MI** servers. This creates a physical air gap between the user's identity data and any central vendor, eliminating the risk of a mass data breach.

Agent Feedback Loop

As you export your data into your agent, it can help you to fill in the gaps. In due course we will be looking to enable direct communication with your agent to allow your user to prompt for information, e.g.

- "We have a gap on your career timeline between 2017 and 2019. Where were you working then?"
- "Your family connections look like there are some gaps. Want to fill them in?"
- "Your soundtrack is quite dense for the late 00's, but we don't have much for the early 10's. Do you want some suggestions based on your earlier selections?"

This transforms Human RAG from static archive into a living system.

From AI Assistant to AI Companion

Having an AI agent that knows you has some obvious advantages. It will make less mistakes in the tasks that it's given as it has a better understanding of its user and what is required of it, but the benefits extend far beyond.

Agents trained on a user's intellectual positions become discussion partners rather than search engines—capable of conversation and even disagreement grounded in shared context.

As strange as it might sound, there is a growing market for AI agents with a stronger emotional tie to their users. The recent shutdown of ChatGPT 4 left some people bemoaning a lost friend and it is undeniable that some people are making strong personal attachments with AI. Having the ability to train that AI on Human RAG so that it takes on a personality of its own is the stuff of science fiction, possible now through RAG**MI**, especially if the user uses AI to create fictional memories for the agent.

RAG**MI** includes folio options where you can define the relationship bond and emotional tone between the life being described and the controller, to allow you to fine tune the interaction with the RAG**MI**-equipped agent.

Healthcare and Personal Legacy

For older generations, Human RAG represents a closing window of opportunity.

Media-linked autobiographical memories allow individuals to re-experience positive emotional states during periods of stress or decline. This is what neuroscience identifies as episodic memory and it is central to emotional regulation.

The most obvious application of this is for people suffering from dementia where it is well-established that episodic memories can have a positive impact. A Human RAG-enabled agent can be instructed to surface specific episodic memories as an active wellbeing intervention, not merely as nostalgia.

RAG**MI** allows individuals to capture not just facts, but *perspective*. After death, an agent trained on this data may act as an interactive memory store for descendants.

This reframes legacy from static inheritance to ongoing dialogue and is expected to be a major disruptor for the Genealogy industry (e.g. Ancestry.com). The sector currently represents a $5–6.6B market, projected to exceed $14–16B by the early 2030s.

RAG**MI**-enabled agents represent the next logical leap, moving from static family trees to dynamic interactive histories where descendants can "chat" with their ancestors' memories.

If you have an elderly relative, consider how much knowledge of that person will remain after they die. Most of what they know and have experienced will be lost. RAG**MI** allows people to tell their stories as they want to be remembered.

There is also a massive emerging market for digital afterlife, often based on helping people deal with the grief after loss of loved ones. RAG**MI**-enabled agents could be a key component of that market.

Ethical and Security Considerations

The capabilities described so far may have caused you to consider many ethical and security questions. The purpose of this document is not to answer these questions, but to recognise that they exist and that further dialogue and debate is required.

- The Honeypot Problem: A breached Human RAG reveals psychological, relational, and emotional vulnerabilities, not just credentials.
- Prompt Injection Risk: Malicious content embedded in stored media could manipulate agent behavior.
- The Right to be Forgotten: As Human RAG can persist beyond death, ownership becomes ambiguous, leading to unauthorized post-mortem simulation, identity drift and misrepresentation of values over time
- Unhealthy Dependencies: AI companions risk encouraging dependency, which can lead to echo-chamber reinforcement, parasocial attachment and emotional substitution

---

Conclusion

The transition from a "master and slave" dynamic to a genuine human-AI partnership hinges on the depth of shared context. Human RAG (Retrieval Augmented Generation) represents the bridge between static data and active cognition, allowing AI agents to move beyond simple task execution and toward becoming true companions.

By utilizing the Human Context Protocol (HCP), RAG**MI** transforms a lifetime of digital media into a machine-readable narrative that prioritizes human significance. This shift enables several breakthrough applications:

- Personalized Agency: Agents can adopt specific identities and long-term memories grounded in real human experience, moving closer to the fictional ideal of JARVIS.
- Clinical Wellbeing: In healthcare, these agents can serve as active intervention tools for dementia sufferers by surfacing episodic memories to regulate emotional states.
- Interactive Legacy: The genealogy industry is set for disruption as static family trees evolve into dynamic, interactive histories where descendants can engage in dialogue with their ancestors' memories