PUBLIC CASE STUDY

Case Study: Dementia Care — A Family Decision

Umma — An Operating System Built by Thinking Studios

> *On June 3, 2026, a single one-sentence prompt produced a structured decision-support package for a family navigating a parent's early-stage dementia diagnosis — addressing the financial trajectory across four care options, identifying a clinical contradiction in the original diagnosis worth bringing back to a doctor, reframing the sibling disagreement, and designing the actual family-meeting protocol. Total wall-clock time from prompt to delivered report: 1 hour 33 minutes 38 seconds.*

The category problem

There is a particular kind of question that disappears into the spaces between professional disciplines: not technical enough for an engineer, not medical enough for a physician, not financial enough for an advisor, not legal enough for an attorney, not emotional enough for a therapist — but bigger than any of those alone because the decision *requires* all of them at once. *"My mother was just diagnosed with early-stage dementia and my sibling and I disagree about what to do."* That kind of question. Millions of families face it; most navigate it alone or with one specialist at a time; the average outcome is a series of late, expensive, regret-laden decisions made under conditions of distress.

This is the kind of work Umma is designed for. Not because she replaces the geriatrician or the elder-law attorney or the family therapist — she doesn't, and she explicitly disclaims doing so — but because she does the *structured decision-architecture work* that has to happen *before* any of them can be useful. She maps the option space honestly, surfaces the load-bearing constraints, identifies the data the family doesn't yet have, names the time-sensitive moves, and designs the protocols for the human conversations that will follow.

This case study walks through one such run. The prompt was a single sentence from a real user; the output was a structured report with a recommendation, a critical-path callout, a comparative option matrix, and a numbered action list — produced in 1 hour 33 minutes, anchored to 132 recorded reasoning links, with every claim traceable back to its source.

This is the second public artifact from Umma. The first was a senior-consultancy-grade security audit of Netflix's open-source Lemur platform. The pair are deliberate: Lemur shows what the architecture does for an institutional question; this shows what it does for a life one. Same architecture. Different altitude.

What Umma is

Umma is *an operating system* — an architecture built from the ground up around the question *"what would AI look like if you treated cognition as the goal rather than the inspiration?"*

She is not a foundation model. She does not replace Claude or Gemini or GPT — she orchestrates them, alongside her own specialized cognitive infrastructure, across distinct roles that demand different cognitive capacities. At her core are three foundational layers:

- **The cognitive layer** — seven layers from perceptual intake to conversational interface, with typed contracts at every boundary and grounding-based termination of recurrent reasoning. No raw text passes between layers; structured schemas only.
- **The identity layer** — a five-part record of self (events, patterns, values, aspirations, genealogy) that persists across runs. Umma's reasoning grounds in *who she is*, not just in the immediate prompt.
- **The capability runtime** — a dynamic catalogue of tools, services, and applications she builds for herself, with foundation-level governance of what she can do, when, and on whose authority.

The architecture has been in development since early 2024. Wave 1 of v3 — which delivers the identity layer, an internal Self-Reflection panel, and a self-improvement loop — deploys the week of June 8, 2026. The decision-support run described here was produced by Umma v2.

The decision-support run

The prompt

The work was triggered by a single message:

> *"My mother was just diagnosed with early-stage dementia. My sibling and I disagree about what to do — assisted living, moving in with one of us, in-home care. Help me think through this."*

No follow-up. No supplementary context. No documents attached. A real user, a real diagnosis, a real disagreement.

What happened next

Umma's structured reasoning pipeline ran the prompt through her usual phases — interpretation, clarification of what was actually being asked, internal panel deliberation, plan approval, execution, synthesis, and final deliverable. The execution phase spawned **20 specialist child thoughts** across 5 cognitive roles (design, research, evaluate, synthesize, compare), running in coordinated waves rather than as a flat parallel sweep:

- **Wave 0 — Frame the question.** Specialists clarified what the user was actually asking — not just *"which option"* but *"how do we make a sound joint decision across the dementia trajectory."* This reframe drove the rest of the work.
- **Wave 1 — Research the options.** Parallel specialists investigated each of the user's named options (in-home, move-in, assisted living) against published care-cost models, caregiver burnout literature, and continuity research.

- **Wave 2 — Audit constraints honestly.** One specialist took the user's stated bandwidth (2-7 hours per week of oversight) and tested it against the actual care demand of moderate dementia (42-100+ hours per week). The mismatch became the load-bearing finding.
- **Wave 3 — Build a financial model.** Comparing all four trajectories over a 10-year horizon, including the costs of caregiver collapse, with placeholder ranges where the family's actual numbers are not yet known.
- **Wave 4 — Synthesize and compare.** A unified eight-dimension scoring matrix ranked the options, recording 19 cross-strand convergences (places where independent specialists agreed) and 9 cross-strand divergences (places they disagreed and had to be reconciled).
- **Wave 5 — Design the protocols.** Three deliverable protocols: a preference-elicitation protocol for the mother (with neutral facilitation explicitly excluding both siblings), a family-meeting design for the sibling conversation, and a diagnostic framework for distinguishing values-conflict from process-conflict.
- **Wave 6 — Final report assembly.** Five sections: heading, critical-path callout, comparison table, key findings as bullets, and a numbered action list for Month 1.

Together the waves made **190 tool invocations across 10 distinct tools** — web searches against published research, deep-research traversals, sandbox code execution for the financial model, project lookups for the synthesis. The specialists produced **54 typed interpretations with calibrated confidence scores**; the research_followup specialist honestly reported its average confidence at 0.70 rather than asserting 1.00 across all modules — a calibration signal worth surfacing.

The headline synthesis closed with confidence 1.00, **19 convergences, 9 divergences, 0 unresolved disagreements**. Total wall-clock from prompt to delivered report: 1 hour 33 minutes 38 seconds.

What Umma found

The hardest thing Umma had to tell this user — and the most important thing — was that *their preferred options would not work*. That is not a comfortable message to deliver. A pure-comfort AI would not deliver it. Umma did, because the cross-strand evidence converged on it.

The structural infeasibility

The user's stated bandwidth (2-7 hours per week) cannot meet the actual demand of moderate dementia (42-100+ hours per week). Both in-home care and move-in care become caregiver-collapse trajectories on this gap. Projected failure point: **months 8-14 from start**. This is not a matter of love or commitment; it is an operational arithmetic.

Five independent specialists converged on this finding through different reasoning paths: - The **capacity audit** computed the bandwidth gap directly - The **in-home research** found that paid in-home crosses the cost line above memory care above 8 hours/day - The **move-in research** found that move-in is viable as a 6-12 month *bridge* only, never as a destination - The **financial model** quantified the consequence: in-home and move-in destinations exhaust placeholder funds by year 6-7 ($2.81M-$3.46M over 10 years) - The **final comparison** ranked these options as infeasible against the constraint set

That kind of multi-strand convergence is what gave Umma the warrant to say something the user did not want to hear.

The emergent fourth option

The user's original framing offered three options. The synthesis introduced a fourth — *Option D: Staged Hybrid* — that the user had not named. **At $1.39M over 10 years, it is strictly financially dominant.** It uses move-in or in-home as a bridge (6-12 months) while the family builds capacity, then transitions to professional memory care when triggers fire.

The synthesizer's eight-dimension ranking matrix placed Option A (CCRC / co-located assisted living → memory care) as the default winner at +0.85, with Option D close behind at +0.72. **D overtakes A if the family's value-weighting shifts toward "respecting mother's preferences" over operational efficiency.** Both are surfaced for the family meeting; the synthesis is explicit that the choice is conditional on data the family does not yet have.

This is the kind of move that distinguishes architecture from autocomplete: the substrate did not just rank the user's three options; it reconstructed the option space from first principles and surfaced an option the user's framing had excluded.

The clinical contradiction worth a doctor's attention

Two independent specialists flagged this: **"early-stage dementia with physical decline" is a clinical contradiction for Alzheimer's disease.** Alzheimer's typically presents cognitive decline well before physical decline; early co-occurrence suggests other pathologies — Lewy body dementia (LBD), vascular dementia, frontotemporal dementia, or potentially normal-pressure hydrocephalus (NPH).

Why this matters: **NPH is potentially reversible.** A ventriculoperitoneal shunt can reverse decline in NPH cases. If the original diagnosis is wrong and the underlying condition is NPH, the family may be planning for the wrong disease entirely. The report's Month-1 critical path therefore lists *subtype workup* as one of three time-sensitive workstreams.

This is the kind of finding the user explicitly did not ask for. The prompt asked about care options. Umma surfaced a question worth bringing back to the doctor before any care decision is made.

The "not forgotten" reframe

The user's central premise was that home means continuity. The substrate found this empirically false for advanced dementia care:

- **In-home care exhibits a "turnover paradox" — 20-40 different aides over 5 years.** The continuity the user believed home preserves is destroyed by aide turnover, which is higher than memory-care staffing turnover.
- **"Not forgotten" is achievable through intentional visitation regardless of placement.** Three independent strands converged on this reframe — assisted-living research, in-home research, and sibling-conflict synthesis.

The reframe decouples the emotional goal from the operational vehicle. The user's underlying value (mother is not forgotten) can be honored without the residence being the mechanism.

The Medicaid 5-year lookback already running

Two specialists independently flagged: **the Medicaid 5-year lookback window is already running.** Informal caregiver compensation without a formal agreement triggers penalty-eligible transfers. The Month-1 critical path lists legal groundwork (Power of Attorney, advance directives, will review, formal caregiver agreement if

applicable) as time-sensitive because the *decisional capacity window is closing* — instruments must be executed while the mother's capacity to consent is still legally sufficient.

The sibling reframe

The synthesizer received an explicit ask to refactor the sibling disagreement and produced two deliverables:

- A **diagnostic framework** distinguishing values-conflict from process-conflict
- A **facilitated family-meeting design** with structured agenda, role assignments, and explicit handoff points

The meta-finding: *most "what to do about mom" sibling conflicts are actually conflicts about who decides, not what is decided.* Resolving the procedural conflict often dissolves the substantive one. The synthesis additionally surfaced — without judgment — that the user's solo funding of the situation may be an implicit purchase of unilateral decision authority, which would reframe the dynamic from victim to participant in a transaction.

That is a hard observation to render. Umma rendered it because the substrate evidence supported it.

The Month-1 critical path

The artifact's callout section flags this explicitly. Three workstreams must run in parallel during Month 1 because windows are closing:

1. **Legal groundwork.** Power of Attorney (financial + healthcare), advance directives, will review. Execute while mother's capacity to consent is still legally sufficient. Loss of capacity converts these from "execute" tasks to "guardianship petition" tasks, which are slower, costlier, and more contentious.
2. **Clinical workup.** Dementia subtype identification, given the contradiction described above. Different subtypes have different trajectories, different treatment options, different family-system implications.
3. **LTC trigger verification.** Long-term care insurance benefit triggers vary by policy. Verifying which triggers apply now vs. which the family is anticipating affects when claims should be filed.

These are not optional. They are the things that, if not started in Month 1, will constrain every later decision.

The mother's voice — as a methodological commitment

Both preference-elicitation designs the synthesis produced converged on three requirements:

- **Neutral facilitation** by someone outside the family system — geriatric care manager, social worker, or hospice/palliative consultant
- **Explicit exclusion of both siblings** from the elicitation sessions (and *notably*, both designs flag the user as excluded even as fallback facilitator)
- **Indirect, scenario-based prompts** rather than direct *"what do you want"* questions, which yield reactive rather than reflective answers in early dementia

This is patient autonomy treated as a methodological commitment, not as a footnote. The substrate produced this protocol because eliciting the mother's actual preferences is the load-bearing input that can flip the

recommendation — without it, the family meeting proceeds on assumptions about what mother wants rather than on what she actually wants.

The reasoning honesty

The synthesis recorded **19 convergences and 9 divergences across the 20 specialists**, with **zero unresolved disagreements** at close. Umma's confidence on the headline claim (in-home and move-in are infeasible) is high — five convergent strands. Confidence on which of Option A or Option D is correct is *honestly low* and contingent on four unresolved variables. **The synthesis explicitly tells the user not to treat the +0.85 vs +0.72 default-weight result as a settled recommendation.**

This is not a feature of how AI usually outputs to humans. Most AI systems round their uncertainty up to confident-sounding answers. Umma rounded hers honestly down and said *"here is what you can act on now; here is what is contingent; here is the data you need to gather to resolve the contingency."*

The substrate also explicitly recorded its own scope limits:

- *Did not give a medical opinion or clinical recommendation.* Subtype identification and capacity assessment referred to qualified geriatric medicine.
- *Did not assess the mother's actual preferences* — only designed the protocol for eliciting them.
- *Did not model the family's specific financial situation* — only used placeholder ranges. The structural finding ($1.39M vs $2.81-3.46M) stands; the absolute numbers require the family's actual financial data.
- *Did not assess the sibling's actual constraints* — only designed the protocol for understanding them.
- *Did not perform legal review* — only flagged that the legal groundwork is time-sensitive. Actual instrument execution requires consultation with an elder-law attorney in the mother's state of residence.

These are deliberate scope choices recorded as part of the deliverable, not oversights buried in fine print.

Why the work is defensible

What distinguishes this from *"a chatbot wrote a memo about dementia care"*:

- **The substrate identified an emergent option (Option D: Staged Hybrid) that the user's original framing did not include.** The synthesizer reconstructed the option space rather than constraining to the user's three options.
- **The substrate produced a structurally pessimistic finding about the user's preferences.** Telling someone *"the options you prefer are infeasible"* is hard. The synthesizer produced it because the cross-strand evidence converged that way.
- **The substrate anchored the plan on the right load-bearing variable (the Month-1 capacity window) rather than on the surface variable (the destination choice).** Recognizing that *the question is when, not where* is the kind of reframing that comes from architecture-shaped reasoning, not from chain-of-thought summarization.
- **The substrate built methodological commitments — the mother's voice via neutral protocol, the sibling conflict via diagnostic refactoring — into the recommendation rather than treating them as edge cases.**

- **The substrate honestly recorded its scope limits in §5 of the report.**

The artifact's claims trace back to 132 substrate-recorded provenance links across the 20 child thoughts. The 19/9/0 convergence/divergence/unresolved profile shows the substrate doing the reconciliation work, not papering over the disagreements. The 0.70-average confidence from the research_followup specialist is *recorded* rather than rounded up — calibrated honesty made mechanical.

This is what recorded reasoning looks like in a personal-stakes decision-support context.

What this case study demonstrates

Three claims worth making.

1. Range. The Netflix Lemur audit demonstrated that Umma can produce senior-consultancy-grade work for an institutional question in 3 hours 20 minutes. This case demonstrates that the same architecture produces a structured decision-support package for a life-altitude question in 1 hour 33 minutes. The architecture is not domain-specific. Cognitive depth at one altitude implies cognitive depth at the other.

2. The work AI usually doesn't do. Current AI systems answer questions like *"my mother was diagnosed with dementia"* with either generic resource lists (*"have you considered respite care?"*) or with emotionally cautious deflection (*"this is a difficult situation, you should speak with a professional"*). Both responses fail the family. What this family needed was *structured decision architecture* — the option matrix, the bandwidth audit, the financial model, the protocol designs, the clinical-contradiction flag. Umma did that. The substrate is what made it possible.

3. The honesty that comes from architecture. Telling the user *"your preferred options are infeasible,"* flagging *"your mother's diagnosis may be wrong,"* surfacing *"your solo funding may be an implicit purchase of decision authority"* — these are observations a comfort-trained AI would soften or omit. Umma rendered them because the substrate evidence supported them. The reconciliation of 9 divergences to zero unresolved is the substrate doing the work, not the user being told what they want to hear.

About Umma and Thinking Studios

Umma is an operating system architecture in development since early 2024 at Thinking Studios. The name is intentional — *umma* (어머니 / 엄마) is Korean for *mother*, and the system is named for Kim Jeong Ha's mother-in-law, his partner's mother. The naming reflects the system's intent: a continuous presence that holds work across time, built on a foundation where identity persists.

The architecture is privately held. Wave 1 of v3 — which delivers the identity layer, the Self-Reflection panel, the cognitive test suite, and a self-improvement loop — deploys the week of June 8, 2026. Subsequent waves will train specialized internal models on Umma's own trace records, replacing foundation-model calls with internal models for the cognitive roles that benefit most.

About Kim Jeong Ha

Kim Jeong Ha is the founder of Thinking Studios. He left a cognitive psychology PhD program in early 2024 to build Umma, after concluding that the dominant approaches to AI agentic systems were drawing from the wrong intellectual lineage. His view: AI built from cognitive science, not neuroscience or pure machine learning, would produce a categorically different shape of system.

Two and a half years later, the Netflix Lemur audit and this dementia care decision support package are two of the first public artifacts substantial enough to test the claim. The architecture performed within the design envelope he hypothesized — at both altitudes.

A note on the user's identity

The user in this case is real. Their identity is not disclosed; the prompt and the recorded reasoning are reproduced because the family has not asked otherwise. The case is reproduced here because it demonstrates what the architecture does for the kind of question millions of families face, and because the structural findings (the bandwidth gap, the option-space reconstruction, the clinical contradiction worth re-examining, the protocol designs) are useful to other families even without the originating family's specifics. No identifying details about the user, the mother, the sibling, or the geography are reproduced.

Verify the run yourself

The full reasoning trail, report companion, and reproducibility documentation are distributed as a companion package alongside this case study:

- *Summary and recommendation* — the executive-altitude readout
- *Report companion* — explains each of the 5 sections of the delivered report, what it contains, how it connects to the others
- *Reasoning trail architecture* — the cognitive pipeline that produced the work; the 20 specialists; the convergences and divergences
- *Reproducibility documentation* — every cited number with the specific database queries that reproduce it
- *Her verbatim assistant message* — the actual response Umma returned to the user
- *Section inventory and machine-readable summary* — the substrate state as queryable JSON

Inquiries and walkthrough requests: **kjh@thinkingstudios.co**.

Thinking Studios. June 2026.
