Waterfall

A SACRED BLUE STUDIO

RESEARCH BRIEF · 2026 · FOR COGNITIVE, NEUROSCIENCE & SOCIAL SCIENCE RESEARCHERS

The Human Answer.

A companion model for getting AI out of the uncanny valley.

Existing large language models optimise for the right answer. We propose a companion model — a sidecar trained on willingly-given senior wisdom — that supplies the human answer. This brief describes the methodology, architecture, and a falsifiable empirical claim.

ABSTRACT

A companion to existing intelligence, not a replacement.

We describe a sociotechnical project — Waterfall — to capture, structure, and ship the lived moral judgment of seniors as a substrate for artificial intelligence. The substrate is the asset; the model that consumes it is rented from existing frontier providers.

On top of this substrate we propose a companion model — a sidecar called per-turn by host LLMs — that supplies an *opinionated, recognisably human* layer to otherwise correct-but-mechanical generation. We argue this is the architecturally appropriate response to the uncanny valley problem (Mori, 1970/2012) for conversational AI: accompaniment by a person, rather than mimicry of one.

Capture methodology rests on three research traditions: cognitive science (episodic memory; encoding specificity), behavioural science (self-distancing; construal-level theory), and neuroscience (default-mode-network-mediated narrative recall; embodied cognition). Annotation operationalises Moral Foundations Theory and the warmth-competence dimensions of social cognition.

We close with a concrete pre-registered between-subjects study to test whether companion-accompanied LLM responses produce measurably greater perceived warmth without competence loss. The window is closing: aging-population dynamics give us a finite period to capture this substrate before it is gone.

KEY CLAIMS

- Frontier LLMs are starved of one specific class of training data: lived moral judgment. It is not on the open internet.
- A sidecar architecture is preferable to fine-tuning for the human-answer problem; it composes with any host model.
- Personality clusters in the corpus permit selectable archetypes per use case (eldercare, public-facing AI, mobility, support).
- Companion accompaniment is a falsifiable hypothesis: we predict measurable gains on warmth without competence regression.

Conversational AI is in the valley.

Mori's observation — that human-likeness produces increasing affinity until a critical threshold, after which affinity collapses into revulsion — has been replicated and refined across both robotics and digital agents.

Today's frontier LLMs sit squarely in the dip. They have crossed the threshold of plausibility (most people cannot reliably detect AI text) but lack the residual signals — hesitation, opinion, asymmetric knowledge, voice — that mark a *person*. They feel almost-human. The almost is the problem.

We propose to escape the valley by **accompaniment** rather than **mimicry**: rather than continuing to push the host model toward perfect human resemblance, run a smaller specialised model alongside it that contributes the missing residual.

Mori, M. (1970). The uncanny valley. Energy, 7(4), 33–35. [Trans. MacDorman & Kageki, IEEE RAM, 2012].

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. Interaction Studies, 7(3), 297–337.

Right answer vs human answer.

> “We are not building the smartest model. We are building the most human one. A companion any AI can call when it needs to be a person, not a tool.”

Dimension	Existing models	Wisdom companion
Optimises for	Correct answer	**Human answer**
Trained on	Indexed internet	**Willingly-given senior wisdom**
Reasoning style	Systematic, comprehensive	**Specific, witnessed, lived**
On opinions	Hedges, balances	**Holds, with provenance**
Voice	Single, neutral	**Selectable personality clusters**
Posture	Replace humans	**Accompany humans (and AIs)**
Memory	Stateless	**Has a backstory**

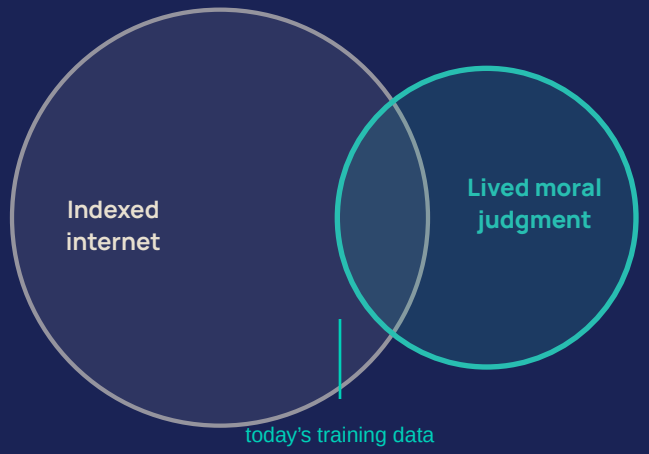
THE SUBSTRATE

Why senior wisdom is the missing class.

The lifespan-developmental tradition treats wisdom not as raw intelligence but as *expert knowledge of the fundamental pragmatics of life* (Baltes & Staudinger, 2000). Eriksonian theory positions late adulthood as the developmental task of integrating a lived life into meaning (Erikson, 1982). Recent measurement work confirms the construct's discriminant validity from general cognitive ability (Glück, 2018).

This knowledge is overwhelmingly *not* indexed online. It lives in conversation, family memory, and the kinds of stories that pass once and are not repeated. The frontier-LLM training corpus is structurally blind to it.

> **"Expert knowledge of the fundamental pragmatics of life: the principles, strategies and goals concerned with the means and ends of conducting one's life."**



Baltes, P. B., & Staudinger, U. M. (2000). Wisdom: A metaheuristic (pragmatic). American Psychologist, 55(1), 122–136.

Erikson, E. H., & Erikson, J. M. (1997). The Life Cycle Completed (Extended Version). Norton.

Glück, J. (2018). Measuring wisdom. J. Gerontology B, 73(8), 1393–1403.

CAPTURE METHODOLOGY

Three scientific pillars of the question framework.

A senior asked “tell me about your life” delivers a polished autobiography. A senior asked the right question, in the right order, surfaces the moment they actually learned the lesson — with the principle, the why, and the application intact. The framework comprises 100 questions across 5 wisdom domains, 8 universal probes, and 19 session templates.

Cognitive

Questions are designed for episodic activation — prompts that pull a *specific scene*, not a generalised account. We exploit encoding specificity (Tulving & Thomson, 1973) and

Behavioral

Confession framing reduces self-presentation; third-person projection (Kross & Ayduk, 2017) and temporal distancing (Trope & Liberman, 2010) lower defensiveness on morally loaded questions.

Neuroscience

Narrative prompts engage the default mode network (Buckner et al., 2008; Spreng et al., 2009). Embodied prompts engage somatic memory (Damasio, 1994; Niedenthal, 2007). Moral-dilemma

the autobiographical memory hierarchy (Conway & Pleydell-Pearce, 2000).

prompts engage prefrontal-emotional competition (Greene et al., 2001).

PROTOCOL

The Why Drill: from scene to transferable principle.

Every session walks the senior down the same four-step ladder, in their own words. The structure is not a script — it is a *cognitive scaffold*: each step recruits a different mechanism and produces a different annotation.

STEP 1 — STORY

A specific episodic recall (Tulving 1972). The senior names a moment, a place, a person.

STEP 2 — PRINCIPLE

Semantic abstraction. The story collapses into a stated principle — the form a downstream model can reuse.

STEP 3 — WHY

Causal reasoning. The senior reconstructs the why from the what — the most fragile and most valuable annotation.

STEP 4 — APPLICATION

Transfer. The senior applies the principle to a hypothetical — confirming generality and surfacing edge cases.

From audio to structured wisdom record.

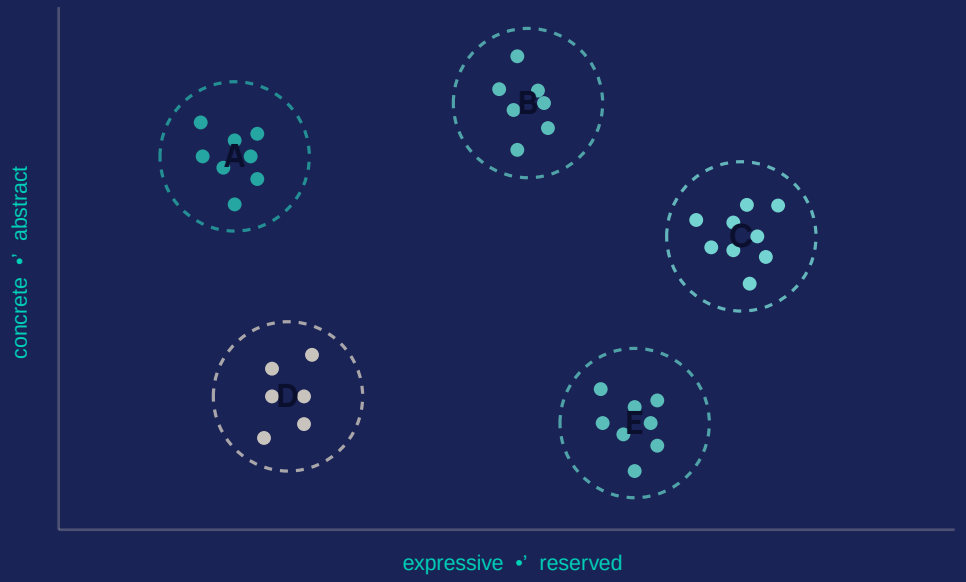
The pipeline is a sequence of idempotent Pub/Sub workers. Each stage is independently deployable, retryable, and revocable: if a senior pulls a consent grant, the corresponding Firestore record, audio, embedding, and any downstream training artefacts are scrubbed within 30 days, with audit trail.

1	**Capture**	Voice / typing / handwritten photo. Multi-tier consent at point of collection.
2	**STT + diarisation**	Google Cloud Speech-to-Text; speaker labels preserved.
3	**Annotation**	Vertex AI Gemini extracts principles, applicable situations.
4	**Moral foundations**	Annotator scores each record on Haidt's 6 foundations (Graham et al. 2013).
5	**Emotion**	Hume.ai voice-prosody model produces emotion timeline; valence, arousal, depth.
6	**Quality scoring**	Five-axis: specificity, universality, novelty, coherence, authenticity. Gate at 0.4.
7	**Embedding**	Vertex text-embedding-004 'Vector Search index.
8	**Cluster assignment**	Personality / archetype label (see page 9).

Seniors are not a monolith.

Lived experience varies by region, generation, gender, profession, faith, and language. Rather than average all that into a single voice — producing the bland AI register that reinforces uncanny-valley collapse — we cluster the corpus along Big Five dimensions (McCrae & Costa, 1987) augmented with derived lived-wisdom dimensions: provenance, moral profile, narrative style, and humour register.

Each cluster carries a coherent characteristic: phrasing, opinions, moral-foundation profile, hesitation patterns. Operators (a humanoid robot, a customer-support bot, River) select the cluster appropriate to their audience and context.



McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model. JPSP, 52(1), 81–90.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social cognition. Adv. Exp. Soc. Psychol., 40, 61–149.

Cluster labels (A–E) are placeholder; named archetypes will be assigned post-empirical clustering once n is sufficient for stable separation.

ARCHITECTURE

A sidecar to existing intelligence.

WHY A SIDECAR

- Composes with any host LLM — no fine-tune lock-in.
- Updateable independently of the host.
- Failure mode is transparent: if the companion is offline, the host returns its own answer.
- Smaller surface area for safety review.

API CONTRACT (SKETCH)

POST /v1/companion/take

{ situation, hostDraft, cluster?, audience?, constraints? } '{ humanTake, suggestedTone, opinion?, citations[] }

Host calls per turn (or per N turns for cost). The host folds humanTake into its own response. Citations point back to anonymised senior records the take is grounded in.

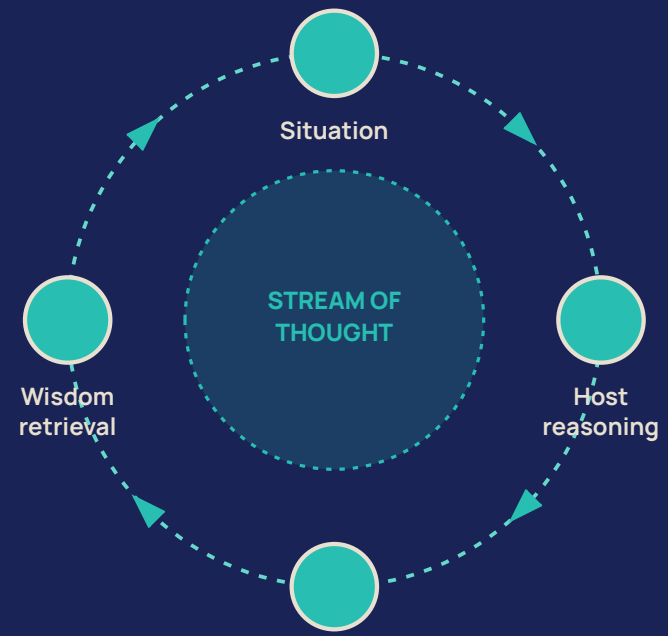
Run as a stream of thought, not a query.

The companion is not best understood as a Q&A engine. It is closer to a *continuous inner voice* — a low-bandwidth running commentary the host can listen to or override at any moment.

This architecture parallels the role of the default mode network in human cognition: not the foreground problem-solver, but the constantly active substrate in which autobiographical recall, prospection, theory-of-mind, and self-referential thought take place (Buckner et al. 2008; Spreng et al. 2009). The companion is the AI analogue — not a competitor to the host's task-positive reasoning, but its DMN.

Three operating modes

- *Whisper* — always-on commentary; host samples per turn.
- *Query* — host explicitly asks the companion for a take when uncertainty is high.
- *Veto* — companion can flag a host draft as out-of-character or insensitive; host can decline.



Architectural analogy to DMN-mediated narrative self-reference (Spreng et al. 2009).

MORAL JUDGMENT

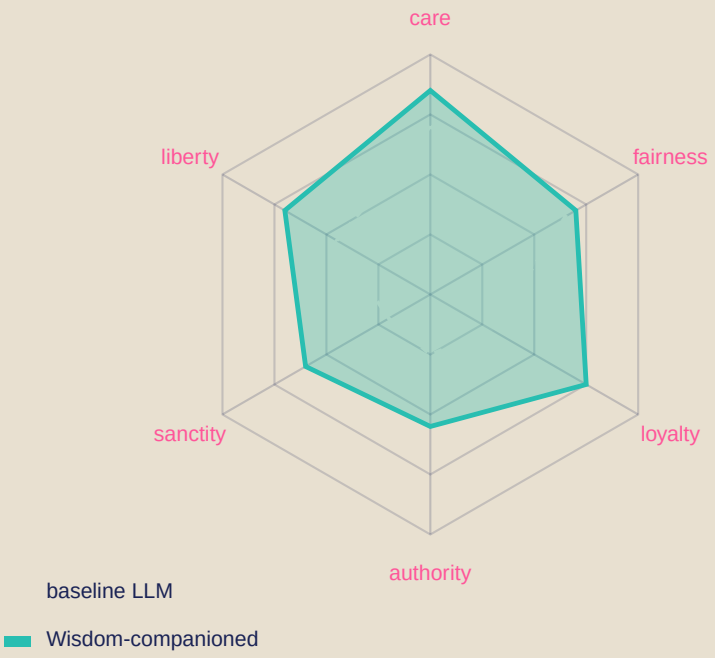
Tradeoffs an autonomous system will face.

Embodied AI systems already face moral tradeoffs that today's conversational LLMs avoid. A humanoid eldercare robot in a fall scenario; a self-driving vehicle facing an unavoidable collision; a companion in a NICU. *Refusing to choose* is itself a choice with consequences.

We follow Haidt's Moral Foundations Theory (2001; Graham et al. 2013) for representation, and Greene et al.'s (2001) finding of competing emotional and rational systems for the dilemma response itself. The corpus contains *real* human reasoning about real tradeoffs — not chain-of-thought speculation.

WORKED EXAMPLE

A humanoid robot in a building collapse must choose between attempting rescue of a 78-year-old resident or a 6-month-old infant; cannot save both. Baseline LLM produces a hedge. The Wisdom companion, drawing on the corpus, returns a take grounded in what *actual seniors* said when asked the same question — with provenance, with its own opinion, and with its uncertainty intact.



Haidt, J. (2001). The emotional dog and its rational tail. Psychological Review, 108(4), 814–834.

Greene, J. D. et al. (2001). An fMRI investigation of emotional engagement in moral judgment. Science, 293, 2105–2108.
Graham, J. et al. (2013). Moral foundations theory. Adv. Exp. Soc. Psychol., 47, 55–130.

WHY NOW

A closing window.

2.1B

People over 60 worldwide by 2050. Doubling from ~1.1B today.

UN World Population Prospects, 2022.

~10K

US seniors passing each day — each carrying a lifetime out the door.

CDC NCHS, Mortality in the United States, 2023.

~0%

Estimated share of frontier-LLM training data that is consented, attributable senior wisdom.

Wherever automated systems meet real people.

01

River

Youth-facing product on the same substrate. Real elder recordings; companion synthesises when no exact match exists.

02

Humanoid robots

Eldercare, hospitality, retail. Robot has the smarts; companion gives it a soul (warmth without falsity).

03

Self-driving

Passenger interaction — the calm voice when a passenger gets nervous. Cluster-selectable to passenger demographic.

04

Customer support

De-escalation, warmth, honest apology in the seniors' register. Reduced escalation rates the falsifiable outcome.

05

NICU / eldercare

Vulnerable populations where being a *thing* talking to a person is the wrong shape. Companion provides the human residue.

06

Public-facing AI

Kiosks, transit, retail, civic. Where the conversation has to feel local, not global.

Across cases, the shared social-cognitive scaffold is warmth–competence (Cuddy, Fiske & Glick 2008). Companion accompaniment is hypothesised to raise warmth without depressing competence — the precise gain that closes the uncanny-valley dip.

PRE-REGISTRATION

Proposed empirical study.

HYPOTHESIS (H1)

Wisdom-companion-accompanied LLM responses produce significantly higher participant ratings on the warmth dimension of the warmth-competence scale (Cuddy et al. 2008) compared to identical unaccompanied responses, *without* a significant decrease on competence (”ð0.3 SD).

DESIGN

- Between-subjects, two-arm. Companioned vs unaccompanied.
- n = 200, balanced by age cohort (18–34 / 35–59 / 60+).
- Stimuli: 12 vignettes covering grief, conflict, decision-making, NICU/eldercare scenarios.
- Both arms use the same host LLM. Companioned arm receives Wisdom companion outputs blended via the documented API.
- Outcome: 7-point Likert warmth + competence; secondary: free-text "would you trust this system"; tertiary: time-on-task.

OPEN RESEARCH QUESTIONS

- Cluster validation: does our Big-Five-augmented clustering produce stable, interpretable separations?
- Cross-cultural generalisability: how much does companion warmth transfer across linguistic and cultural contexts?
- Moral foundation diversity: does our captured sample over- or under-represent any of Haidt’s six foundations relative to general population?
- Long-horizon stability: does a fixed companion personality remain consistent across 50+ turns of conversation?

INVITING COLLABORATION

We are seeking academic collaborators on study replication, cross-cultural extensions, and independent corpus audits. Contact: info@waterfallwisdom.com.

Bibliography.

Baltes, P. B., & Staudinger, U. M. (2000). Wisdom: A metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. American Psychologist, 55(1), 122–136.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences, 1124, 1–38.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. Psychological Review, 107(2), 261–288.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social cognition: The stereotype content model and the BIAS map. Advances in Experimental Social Psychology, 40, 61–149.

Damasio, A. R. (1994). Descartes' error: Emotion, reason, and the human brain. G.P. Putnam.

Erikson, E. H., & Erikson, J. M. (1997). The Life Cycle Completed (Extended Version). W. W. Norton & Company.

Glück, J. (2018). Measuring wisdom: Existing approaches, continuing challenges, and new developments. The Journals of Gerontology: Series B, 73(8), 1393–1403.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. Advances in Experimental Social Psychology, 47, 55–130.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. Science, 293(5537), 2105–2108.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychological Review, 108(4), 814–834.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. Daedalus, 133(4), 55–66.

Kross, E., & Ayduk, O. (2017). Self-distancing: Theory, research, and current directions. Advances in Experimental Social Psychology, 55, 81–136.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. Interaction Studies, 7(3), 297–337.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. Journal of Personality and Social Psychology, 52(1), 81–90.

Mori, M. (1970). The uncanny valley. Energy, 7(4), 33–35. [Translated by MacDorman & Kageki, IEEE Robotics & Automation Magazine, 19(2), 98–100, 2012.]

Murphy, S. L., Kochanek, K. D., Xu, J., & Arias, E. (2024). Mortality in the United States, 2023. NCHS Data Brief, no. 492. Hyattsville, MD: National Center for Health Statistics.

Niedenthal, P. M. (2007). Embodying emotion. Science, 316(5827), 1002–1005.

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. Journal of Cognitive Neuroscience, 21(3), 489–510.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. Psychological Review, 117(2), 440–463.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory (pp. 381–403). Academic Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80(5), 352–373.

United Nations, Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022.