

HARMONIA

TASK ASSIGNMENT

Task	PIIP Personality Cluster Generation via Claude Agent
Issued By	Avery — Operations Manager
Date Issued	10 February 2026
Priority	High — Required for beta personality database population
Assigned To	Abe (Lead), David (Co-developer), Joles (Oversight), Augustine (Technical QA)

1. Context & Background

This month, I (Avery) will have limited availability for Harmonia engine work due to university priorities and a mock exam on the 28th. To keep development on track, I am delegating structured weekly tasks to the team. **This is purely a scheduling constraint** and has nothing to do with anyone's performance.

The PIIP (Personality Inference through Interactive Prompting) system requires a populated database of diverse personality clusters to function at beta launch. Our 6-question scenario-based assessment needs training data — specifically, a wide range of human-like responses that express the full spectrum of Seven Deadly Sins trait combinations. Manually writing these would be prohibitively slow, so we need an automated solution.

2. Task Objective

Build a Claude-powered agent that automatically generates diverse, human-like responses to the 6 PIIP scenario questions, producing personality clusters that can be stored in Abe's database and later used for training and matching.

2.1 How This Fits the Pipeline

Important: The Claude Haiku agent is only responsible for **generating** human-like text responses. The responses are then submitted to the existing PIIP pipeline, where **Gemini AI** (not Claude) parses them into Seven Deadly Sins trait scores. The pipeline is: Claude Haiku generates responses → responses submitted to POST /api/questionnaire/submit → Gemini parses into sin scores → profile stored in database. Do not confuse the two models' roles.

2.2 Success Criteria

1. The agent generates responses to all 6 questions in a single run
2. Each response is 50–100 words and reads naturally (not robotic or formulaic)

- Responses show multifaceted human behaviour — partial expressions of multiple sins per answer, not single-trait caricatures
- The agent can run continuously in a loop, generating unique personality profiles each time
- Generated responses are stored directly into the PIIP database schema
- Cost per profile remains minimal (Haiku model selected for token efficiency)

3. The 6 PIIP Questions

Each question targets a distinct life domain and is designed to activate multiple personality dimensions simultaneously. The sin mappings below are taken directly from the PIIP specification (Appendix B.1). The agent must respond to all 6:

#	Question	Domain Tested	Primary Sins	Secondary Sins
Q1	The Group Dinner Check	Resource sharing, social dynamics	Greed, Wrath	Pride, Sloth
Q2	The Unexpected Expense	Stress response, financial reasoning	Greed, Sloth	Wrath, Gluttony
Q3	The Weekend Off	Leisure priorities, energy restoration	Sloth, Lust, Gluttony	Pride
Q4	The Unequal Split	Conflict handling, fairness expectations	Wrath, Pride	Envy, Sloth
Q5	The Friend Crisis	Loyalty trade-offs, relationship prioritisation	Lust, Wrath	Pride, Greed
Q6	The Feedback Received	Ego resilience, growth orientation	Pride, Wrath	Envy

4. Claude Agent Specification

4.1 Model Selection

Model: Claude Haiku 4.5 (claude-haiku-4-5) — selected as the cheapest available Claude model to minimise token spend during high-volume generation. The task does not require deep reasoning; it requires fast, varied, human-like text at scale. As of February 2026, Haiku 4.5 is priced at \$1 per million input tokens and \$5 per million output tokens.

Use the model string **claude-haiku-4-5** (this always points to the latest Haiku 4.5 snapshot). The dated version is `claude-haiku-4-5-20251001`. Refer to the Anthropic API documentation at <https://docs.claude.com> to verify the model string is still current before implementation, and for rate limits and any pricing changes.

4.2 Prompt Design Philosophy

The agent must not simply “answer the questions.” It must simulate the range of ways real humans would respond. This means the prompt must instruct the model to:

- Adopt a randomised personality archetype before answering (e.g., “anxious people-pleaser,” “confident pragmatist,” “laid-back avoider”)
- Write in first person, casual tone, as if texting a friend or writing in a journal
- Show mixed motivations in each answer — not pure virtue or pure vice, but the messy blend real people exhibit
- Naturally and partially surface behaviours that map to the Seven Deadly Sins framework (greed, pride, lust, wrath, gluttony, envy, sloth) without being explicitly designed to always target these
- Vary sentence structure, vocabulary level, and emotional tone between runs

Reference: Read the “*Philosopher’s Take on Splitting the Bill*” example for the tone and depth we’re targeting. The responses should feel like a real person thinking through a scenario, not an AI summarising traits.

4.3 Recommended Prompt Template

Below is a starting template. Abe and David should refine this during implementation:

SYSTEM PROMPT (draft):

You are a personality response simulator. For each scenario question, you will:

1. First, silently generate a random personality archetype (age, temperament, values, communication style). Do NOT include this in your response.
2. Answer each question in 50-100 words as that person would, in first person, casual written tone.
3. Show realistic, mixed motivations. Real people are contradictory — someone can be generous but also quietly resentful.
4. Do NOT label traits, sins, or personality types in your response.
5. Vary your style each time — different vocabulary, emotional tone, sentence structure.
6. Return ONLY a JSON object with keys q1 through q6.

4.4 Output Format

Each agent call should return a JSON object that maps directly to the questionnaire submission endpoint:

```
{
  "q1": "Honestly I just suggest splitting evenly...",
  "q2": "First thing I'd do is check my savings...",
  "q3": "Probably sleep in till noon then..."
```

```
"q4": "I'd be annoyed but I'd probably...",
"q5": "Oh man, that's a tough one...",
"q6": "I'd need a minute to process it..."
}
```

This JSON output is then submitted to the **POST /api/questionnaire/submit** endpoint, which triggers **Gemini** (not Claude) to parse the responses into Seven Deadly Sins trait scores, as documented in the PIIP specification.

4.5 Loop Architecture

The agent runs in a continuous loop:

7. Call Claude Haiku API with the system prompt + all 6 questions
8. Parse JSON response
9. Validate each response is 25–150 words (the PIIP design constraint; the API hard limit is 250 but the intended range is 25–150)
10. Submit to the questionnaire endpoint with a generated synthetic user_id
11. Log success/failure
12. Wait a brief interval (to respect rate limits), then repeat

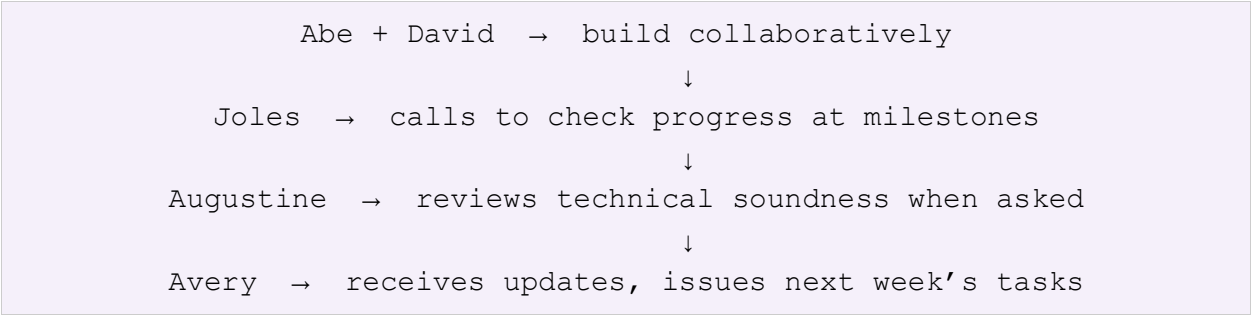
Target: Generate as many diverse personality clusters as possible within budget constraints. Aim for 100+ profiles in the first sprint.

5. Management & Accountability Structure

5.1 Roles

Person	Role	Responsibilities
Abe	Lead Developer	Coordinates with David on architecture and implementation. Owns the database integration. Primary point of contact for technical decisions.
David	Co-Developer	Works collaboratively with Abe on building and testing the Claude agent. Shares implementation responsibility.
Joles	Oversight / PM	Calls Abe or David to check progress. When a milestone is reached, escalates to Augustine for technical QA. Keeps Avery updated.
Augustine	Technical QA	Reviews the Claude bot's technical soundness when Joles escalates. Validates prompt quality, output format, and database integration. This is a lightweight review role, not heavy development.
Avery	Delegator	Issues weekly task assignments. Available for questions but not actively developing this month.

5.2 Communication Flow



6. Milestones & Checkpoints

Joles should check in with Abe/David at each of these milestones and escalate to Augustine where noted:

#	Milestone	What to Check	Escalate to Augustine?
M1	API connection working	Can call Claude Haiku and get a text response	No
M2	Prompt producing human-like responses	Read 5+ sample outputs — do they sound like real people?	Yes — review prompt quality
M3	JSON output valid and parseable	Output matches the q1–q6 JSON schema	Yes — review format
M4	Database integration working	Generated profiles appear in the PIIP database correctly. Confirm Gemini parses them successfully.	Yes — review DB writes
M5	Loop running and producing diverse output	Run 20+ iterations — check for repetition or patterns	Yes — final review

7. Cost Considerations

Why Haiku: We are using Claude Haiku 4.5 because this task is high-volume, low-complexity text generation. We do not need deep reasoning — we need speed and cost efficiency. Each profile consists of 6 responses at 50–100 words each, which is a small token footprint. At current pricing (\$1/MTok input, \$5/MTok output), a single profile generation costs well under \$0.01. Generating 1,000 profiles should cost roughly \$5–10 total. Refer to <https://docs.claude.com> for the latest pricing details.

Batch API: If generating large volumes, consider using the Anthropic Batch API, which offers a 50% discount on token costs for non-urgent workloads processed within 24 hours. This could halve the generation cost.

Important: Abe and David should check the current Anthropic API documentation for rate limits before implementation. Haiku 4.5 rate limits vary by usage tier and may throttle high-volume loops.

8. Important Notes

- **Two models, two roles.** Claude Haiku generates the synthetic responses. Gemini AI parses those responses into Seven Deadly Sins scores. These are separate steps — do not try to have Claude do the sin scoring, and do not try to have Gemini generate the responses. Each model is chosen for its specific strength in the pipeline.
- **The agent is a data generator, not a replacement for real users.** These synthetic profiles seed the database so the matching algorithm has training data. They will be supplemented (and eventually replaced) by real beta tester data.
- **Quality over quantity.** If the responses are repetitive or formulaic, the personality clusters will be useless. Spend time refining the prompt before scaling up the loop.
- **Augustine's role is lightweight.** Augustine is helping Joles with technical review, not building the system. Please respect this boundary — it should not require significant time commitment.
- **Word count enforcement.** The PIIP design specifies 25–150 words per response (Section 2.6 of the PIIP spec). The API hard limit is 250 words (MAX_WORD_COUNT config), but the intended design range is 25–150. The agent's prompt should target 50–100 words, and validation should reject responses outside the 25–150 range to stay consistent with the design intent.
- **Mark synthetic data.** All generated profiles should be clearly flagged as synthetic in the database (e.g., a "source: claude_agent" field) so they can be distinguished from real user data.

9. Reference Document

The complete PIIP specification is documented in **ilovepdf_merged.pdf**, which covers every aspect of the personality system in detail: the question design rationale, Gemini parsing architecture, scoring mechanics, profile aggregation pipeline, similarity calculation, deployment configuration, error handling, testing strategy, and worked examples with full scoring breakdowns.

If anything in this task assignment is unclear or if you are unsure how a component should work, defer to ilovepdf_merged.pdf. It is the authoritative source of truth for the entire PIIP system. The document includes complete prompt templates, JSON schemas, API endpoint specifications, and code implementations that should answer any technical questions about how the pipeline functions.

Questions? Reach out to Avery. Let's keep building.

Harmonia — Chemistry, Not Selection