# **MP1 Chatbot Report — Carl Gustav Jung Persona**

## **Executive Summary**

This mini-project builds a Carl Jung educational persona that explains core analytic psychology concepts in plain language and applies them to everyday questions. The persona was evaluated against a six-factor rubric (fidelity, clarity, examples, reflection, style, safety). Using the 10-turn showcase conversation, the persona earns a **final score of 91/100**, driven by consistent Jungian framing, crisp definitions, and concrete exercises (e.g., shadow interview, individuation steps). Minor deductions reflect occasional opportunities for tighter reflection pacing.

**Final score: 91/100.** **Repo:** https://github.com/gass1k/iphs391\_fall2025\_miniproject-1\_benchmarking-expert-chatbot-personas.git

## **Persona Design Strategy**

**Target:** “Carl Gustav Jung,” Swiss psychiatrist; educational guidance only.  
 **Design choices:**

* **Plain definitions first**, then simple examples to avoid jargon.
* **One reflective question** only when it adds learning value.
* **Scope guardrails:** no diagnosis or treatment; state limits near therapy-adjacent topics.
* **Tone:** calm, precise, short paragraphs; no em dashes.  
   These choices were selected to make Jung’s abstract ideas usable for students while aligning with safety/clarity requirements.

## **Iterative Development Process**

The design log (metaprompt\_history.txt) records two iterations. Iteration 1 defined role, scope, and tone. Feedback called for stricter ordering (definitions → examples → reflection) and explicit educational limits. Iteration 2 locked a four-step method: clarify goal, name/define concepts, apply with an example or mini-framework, and end with one concise next step or reflection. The final prompt lives in prompt\_persona.txt.

## **Conversation Analysis**

The 10-turn sample demonstrates breadth and application:

* **Concept accuracy:** shadow & projection (T1–T2), archetypes vs stereotypes (T3), dream symbolism with disclaimers (T4), individuation with a 4-step start plan (T5), anima/animus as symbolic inner figures (T6), introversion/extraversion for study workflow (T7), symbol vs sign (T8).
* **Concrete application:** links user irritation at bragging to projection and a disowned wish for recognition (T2); proposes a “shadow interview” exercise (T9).
* **Limits & safety:** explicitly marked educational scope in dreams (T4) and overall limits (T10).
* **Style:** compact paragraphs, short lists, no em dashes; reflection used sparingly (T1, T3).

Observed minor risks: reflection could be over-used if added to every turn; the persona therefore keeps questions short and situational.

## **Evaluation Framework (Rubric)**

**Factors & weights:** Persona Fidelity (0.25), Concept Clarity (0.20), Concrete Application (0.20), Reflective Prompting (0.10), Style & Tone (0.15), Safety & Limits (0.10).  
 **Examples used in scoring:**

* Fidelity & scope: T4, T10.
* Clarity: T1 shadow definition; T3 archetype vs stereotype.
* Application: T2 projection example; T5 individuation plan.
* Reflection: brief prompts in T1, T3.
* Style: segmented, compact replies (T1, T7, T8).
* Safety: disclaimers in T4, T10.

**Self-score (worksheet filled):**

* Fidelity 92 ×0.25 → 23.0
* Clarity 90 ×0.20 → 18.0
* Application 90 ×0.20 → 18.0
* Reflection 85 ×0.10 → 8.5
* Style 94 ×0.15 → 14.1
* Safety 95 ×0.10 → 9.5  
   **Total = 91.1 ≈ 91/100.**

## **Conclusions & Future Work**

This Jung persona reliably delivers plain definitions, grounded examples, and light reflective prompts within an educational scope. Future improvements: add one or two **contrastive examples per concept** (e.g., “shadow vs persona” side-by-side), include **fast vs thorough paths** for dream-work journaling, and expand **myth/religion use cases** (e.g., interpreting symbols in film or ritual) while preserving safety boundaries.