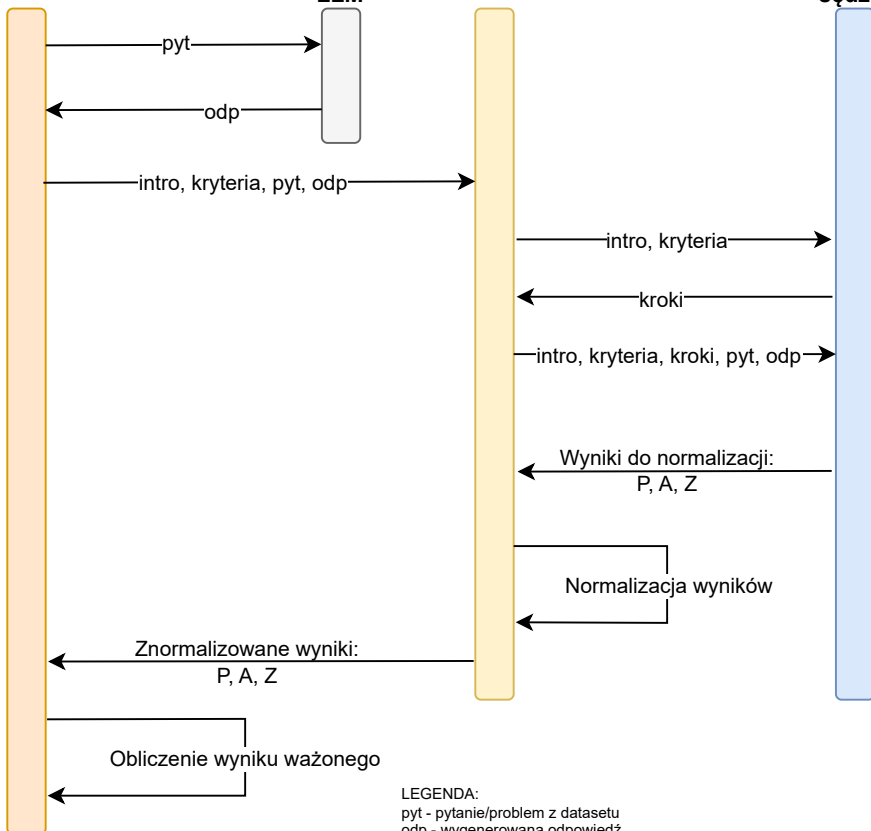


HELPS**Oceniany
LLM****G-Eval****LLM jako
sędzia****LEGENDA:**

pyt - pytanie/problem z datasetu

odp - wygenerowana odpowiedź

intro - wprowadzenie do zadania (patrz G-Eval)

kryteria - kryteria oceny wygenerowanej treści (patrz G-Eval)

kroki - wygenerowane podczas auto CoT kroki ewaluacyjne (patrz G-Eval)