

Od polecenia do agenta

Najniższy poziom, który dowodzi · 15 min

Max Małecki · Tech Lead @ Insly

Dobre polecenie ma cztery części

Tak nie:

Napisz coś o pogodzie dla turystów.

Tak tak:

Cel: ostrzeżenie pogodowe dla wypożyczalni kajaków nad Śniardwami.

Kontekst: prognoza poniżej. Odbiorca to turysta bez doświadczenia żeglarskiego.

Ograniczenia: maksymalnie 3 zdania, bez żargonu meteorologicznego.

Rezultat: zdanie o zagrożeniu, zdanie co robić, zdanie kiedy sprawdzić ponownie.

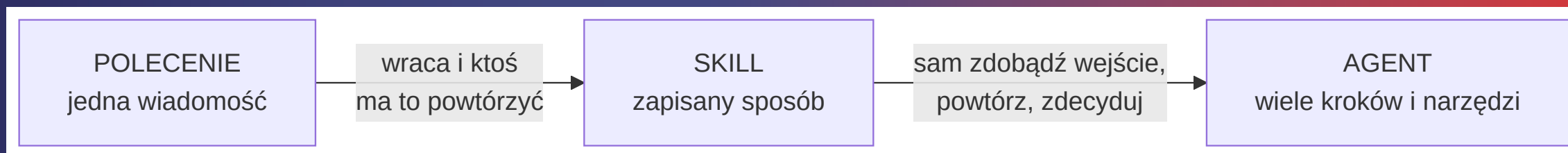
Wynik poprawia się w pętli, nie za pierwszym razem

Uruchom, oceń, popraw, przetestuj ponownie.

- Kryterium oceny ustalacie **przed, nie po**. Inaczej „lepiej” znaczy „bardziej mi się podoba”
- Zmieniajcie **jedną rzecz naraz**, bo inaczej nie wiecie, co pomogło
- Ten sam wsad za każdym razem. Bez tego niczego nie porównujecie

Trzy iteracje na dobrym przykładzie biją trzydzieści na przypadkowych.

Trzy poziomy tego samego zadania



Każdy krok w prawo to więcej mocy i **więcej rzeczy, które się psują**.

Poziom 1: polecenie

Wklejacie prognozę, dostajecie jedno ostrzeżenie.

Cel: ostrzeżenie pogodowe dla wypożyczalni kajaków nad Śniardwami.

Kontekst: [wklejona prognoza]. Odbiorca to turysta bez doświadczenia.

Ograniczenia: maksymalnie 3 zdania, bez żargonu.

Rezultat: zagrożenie, co robić, kiedy sprawdzić ponownie.

Kiedy to wystarcza: robicie to raz, dla jednego miejsca, i sami czytacie wynik.

Poziom 2: skill

Skill to zapisany, nazwany sposób wykonania zadania. Ten sam prompt, który już sprawdziliście, wyjęty z czatu i zapisany tak, żeby dało się go użyć ponownie i oddać komuś innemu.

- Miejsce na wejście: `{{prognoza}}`, `{{nazwa_akwenu}}`
- Kształt wyjścia zabetonowany. Trzy zdania, zawsze te same trzy
- Każdy z zespołu uruchamia identycznie, więc **wyniki da się porównać**

Kiedy to wystarcza: ten sam kształt zadania wraca, a robi to więcej niż jedna osoba.

Poziom 3: agent

Ten sam przykład, ale nikt już nic nie wkleja.

- Sam pobiera prognozę, bo ma dostęp do sieci i do plików
- Przechodzi **12 lokalizacji** zamiast jednej
- Zapisuje plik z gotowymi ostrzeżeniami
- Oznacza te, które przed publikacją wymagają akceptacji człowieka

Hermes czyta i zmienia Wasze pliki oraz uruchamia komendy na Waszych uprawnieniach. Dajcie mu katalog projektu, nie cały dysk.

Kiedy który — trzy warunki

Zadajcie je o **swoim** zadaniu, w tej kolejności:

1. **Jedno wejście, jedno wyjście, robicie to raz?** Zostańcie przy poleceniu.
2. **Ten sam kształt wraca i ktoś inny ma to powtórzyć tak samo?** Zapiszcie skill.
3. **Trzeba samemu zdobyć wejście, powtórzyć na wielu elementach i zdecydować, co dalej?** Dopiero to uzasadnia agenta.

Jeśli odpowiedź na trójkę brzmi „nie”, nie budujcie agenta. Serio.

Co się psuje na każdym poziomie

Poziom	Co się psuje
Polecenie	nie da się powtórzyć, każdy dostaje inny wynik
Skill	sztywnieje, przy nietypowym wejściu wychodzi bzdura w dobrym formacie
Agent	koszt i nieprzewidywalność, a przy awarii nie wiadomo, który krok zawiódł

U nas w Insly **120 000 PDF-ów miesięcznie obsługuje workflow, nie agent.**

Kroki są stałe, więc agent dołożyłby tylko koszt i nieprzewidywalność.

Agent, który halucynuje, jest gorszy niż formularz, który działa

Wybierzcie najniższy poziom, który dowozi.