L O G I , P R O J E K T Y , S E S J E

Jak KreionAI zarządza życiem agentów

Model pracy, który wynika z bezstratnej pamięci BTTF

Mirosław Kardaś | Atnel

NAWYK, KTÓRY TRZEBA ODUCZYĆ

Każde narzędzie AI, jakiego używałeś do tej pory, nauczyło Cię jednego równania: czat = kontekst. Nowa rozmowa to nowy agent, stara rozmowa to stary agent, a wszystko, co agent wie, mieszka w tej jednej rozmowie. To równanie jest tak powszechne, że nikt go nawet nie wypowiada - po prostu tak działa świat.

I z tego równania wynika cała patologia dzisiejszej pracy z agentami. Skoro wiedza mieszka w sesji, to sesja jest skarbem: nie wolno jej zamknąć, bo wszystko przepadnie. Więc użytkownicy hodują sesje tygodniami, konteksty puchną, model coraz słabiej widzi igłę w stogu, aż w końcu przychodzi kompaktacja - automatyczne streszczenie, które po cichu wycina szczegóły - i człowiek odkrywa, że jego "skarb" został podmieniony na notatkę z tego skarbu. A narzędzia terminalowe dokładają drugą skrajność: każde uruchomienie to nowa sesja, więc na dysku rośnie cmentarzysko setek kontekstów bez nazw, do których nikt nigdy nie wróci, bo nie sposób ich odnaleźć.

KreionAI zrywa to równanie u samych podstaw. I to jest pierwsza rzecz, którą trzeba zrozumieć, żeby poczuć, jak się w nim pracuje.

TRZY BYTY ZAMIAST JEDNEGO ZROSTU

Zanim padną definicje - jedno rozstrzygnięcie terminologiczne, od którego zależy zrozumienie całej reszty. W każdym narzędziu poza Kreionem "sesja" oznacza rozmowę - i prawie nikt nie zdaje sobie sprawy, że tam ta rozmowa jest w rzeczywistości odpowiednikiem 1:1 fizycznego kontekstu agenta. Jeden czat = jeden kontekst: dwa słowa na ten sam byt, zrośnięte tak szczelnie, że nikt na świecie tego nie prostuje, bo dopóki zrost istnieje, rozróżnienie jest po prostu niepotrzebne. Kreion ten zrost rozcina - i dlatego potrzebuje dwóch osobnych słów tam, gdzie cały świat radził sobie jednym. Rozmowa to u nas LOG. A słowo "sesja" znaczy w tym tekście zawsze i wyłącznie jedno: **sesja agenta, czyli jego fizyczny plik kontekstu** - realny byt na dysku, w którym żyje chwilowa pamięć robocza agenta. Jeśli czytając dalej choć raz złapiesz się na myśli "sesja, czyli rozmowa" - wróć do tego akapitu, bo to nawyk wyniesiony z narzędzi, w których tych dwóch pojęć nigdy nie trzeba było rozdzielać.

W Kreionie to, co wszędzie indziej jest jednym obiektem, zostało rozdzielone na trzy niezależne byty o różnych cyklach życia.

Log to czysty zapis rozmowy - i nic więcej. Nie jest kontekstem, nie jest agentem, jest dokumentem. Jako składnik kartridża BTTF nie umiera nigdy: każde słowo, każda decyzja i każdy załącznik zostają w nim na zawsze, do dyspozycji każdego przyszłego agenta.

Sesja to fizyczny kontekst agenta - jego pamięć podręczna, jego "tu i teraz". Jest tania, jednorazowa i całkowicie wymienna. Można ją zakończyć jednym kliknięciem i nic wartościowego przy tym nie ginie, bo wartość nigdy w niej nie mieszkała.

Projekt to właściciel całości: spina zapis, kontekst i sposób pracy w jedną trwałą tożsamość.

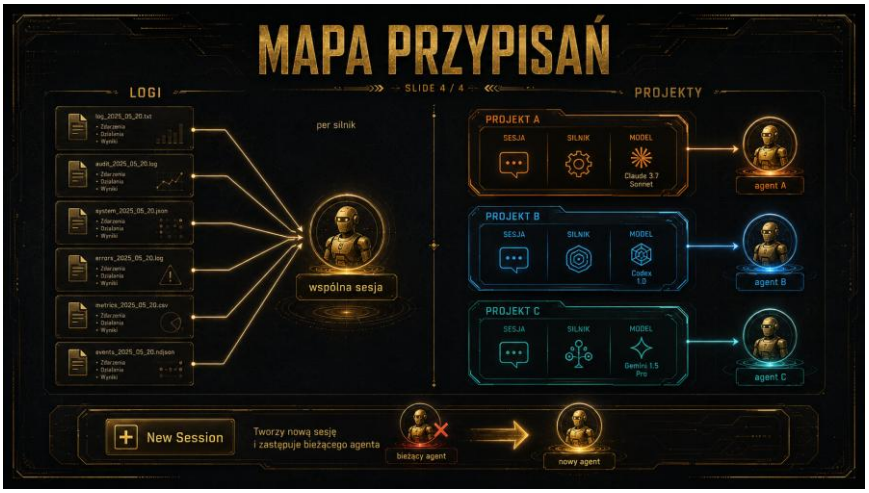
Inżynierom systemów ta separacja powinna wyglądać znajomo, bo to jest czysty podział ról znany z architektury komputerów: kartridż to pamięć trwała (***persistent storage***), log to niezmienny zapis historii (***immutable history***), sesja to zbiór roboczy (***working set***), a model to wymienny procesor (***replaceable processor***). I z tego podziału wynika zmiana kategorii całych klas problemów: utrata sesji przestaje być katastrofą, a staje się zwykłym ***cache miss*** - sytuacją znaną z każdego procesora: danych chwilowo nie ma pod ręką, więc system pobiera je ponownie z pamięci trwałej i pracuje dalej, bo w pamięci podręcznej nigdy nie było niczego, co istniałoby tylko tam; zmiana modelu przestaje być migracją pamięci, a staje się wymianą procesora; przeniesienie pracy na inny komputer przestaje być rekonstrukcją świata z kilku usług, a staje się uruchomieniem tego samego stanu trwałego w nowym miejscu.

Rozdzielenie tych trzech rzeczy jest możliwe tylko dzięki BTTF. Sesję wolno traktować jak materiał zużywalny wyłącznie wtedy, gdy wiedza nie mieszka w sesji - a w Kreionie nie mieszka. Sejfem jest kartridż. Sesja jest rękawiczką.

REGUŁY GRY: LOGI

Zwykłe logi - te tworzone przyciskiem New chat - żyją wszystkie na jednej wspólnej sesji agenta. Nowy czat NIE oznacza nowego agenta: to ten sam rozmówca, z tym samym "tu i teraz", tylko zapis idzie do nowego dokumentu. Również uruchomienie aplikacji nie powołuje nowego agenta - Kreion wraca do tego, z którym pracowałeś, jakbyś nigdy nie wychodził.

Stąd dwie proste zasady praktyczne. Po pierwsze: logi służą do krótkich, bieżących spraw - szybkie pytanie, drobna analiza, pojedynczy temat. Po drugie: jeśli wracasz do rozmów po dniu czy dwóch i zaczynasz nowy wątek, zacznij od **New Session** - zakończ pracę dotychczasowego agenta i powołaj świeżego. To jest świadoma decyzja użytkownika, nie automat: w Kreionie sesje nie rozmnażają się same przy każdym uruchomieniu ani nie ciągną się w nieskończoność przez przypadek. Powstają wtedy, kiedy Ty o tym decydujesz - i dlatego jest ich mało, a każda ma sens.



Mapa przypisań - logi na wspólnej sesji agenta, projekty z własną sesją, silnikiem i modelem

REGUŁY GRY: PROJEKTY

Dowolny log można jednym ruchem przerobić na projekt - i wtedy zaczyna się to, czego nie ma nigdzie indziej.

Projekt dostaje własną, przypisaną na stałe szufladkę, w której trzyma komplet swojego stanu: swoją sesję, swój silnik i swój model. Klikasz w projekt - i jesteś z powrotem u SWOJEGO agenta: ten sam kontekst, ten sam dostawca, ten sam model, dokładnie tam, gdzie skończyliście. Klikasz w drugi projekt - teleportujesz się do innej szufladki, w której może pracować agent zupełnie innej firmy. Projekt pamięta nawet to, który silnik nad nim pracuje: jeden projekt może żyć u Claude'a, drugi u Codexa, a przełączanie się między nimi to jedno kliknięcie w pasku bocznym - bez żadnej gimnastyki z wyszukiwaniem sesji, nadawaniem im nazw i zgadywaniem, która była która, znanej każdemu użytkownikowi narzędzi terminalowych.

Dobra praktyka na start: tworząc projekt, kliknij New Session - świeżo powołana sesja zostanie przypisana do projektu i będzie jego domem.

Ale uwaga - szufladka nie jest po to, żeby jedną sesję hodować miesiącami. W projektach obowiązuje ta sama higiena co wszędzie: konteksty zmienia się jak rękawiczki. Skończyłeś etap, zamykasz temat, rano ruszasz z czymś nowym? Zakończ jego sesję i pracuj na świeżym agencie. Kompaktacji się nie hoduje - dopuszcza się ją świadomie i rzadko, właściwie tylko wtedy, gdy pracujesz nad planem tak dużym, że nie mieści się nawet w milionowym oknie kontekstu. I nawet wtedy nie ma czego się bać: nad każdą sesją czuwa kartridż, więc jeśli agentowi po kompaktacji coś się "rozmyje", jednym poleceniem doprowadzasz go z powrotem do formy.

ŚWIEŻY AGENT TO NIE STRATA. TO JAKOŚĆ

W narzędziach sesyjnych świeża sesja to katastrofa: wszystko trzeba tłumaczyć od nowa, więc ludzie robią wszystko, żeby jej uniknąć - i tym samym hodują degradację własnego kontekstu. Kreion odwraca znak tego równania.

Świeżo powołany agent ma czysty kontekst, pełną uwagę modelu i zero szumu po poprzednich tematach. A wiedzę zaciąga sam: mówisz mu "zbriefuj się z tematu X" - i nieważne, w którym logu czy projekcie ten temat leżał - agent rusza w kartridż jak chart za zdobyczą, przynosi komplet ustaleń i pracuje dalej, jakby był z Tobą od początku. Świeży kontekst nie jest amnezją - jest odśmiecaniem poznania (***garbage collection*** dla ***cognition***): usuwa szum, a dostęp do wszystkiego zostawia nietknięty. Świeżość przestaje być kosztem, a staje się przewagą: najlepiej pracuje się z agentem, który ma czystą głowę i pełny dostęp do archiwum.

W informatyce od lat funkcjonuje rozróżnienie, które pasuje tu idealnie: zwierzątka i bydło (***pets vs cattle***). Serwerom-zwierzątkom nadaje się imiona, leczy się je i opłakuje; bydło się po prostu wymienia, bo wartość jest gdzie indziej. Konkurencja hoduje sesje jak zwierzątka - bo musi, bo bez nich traci wszystko. Kreion może traktować sesje jak bydło - bo wszystko, co cenne, mieszka w kartridżu. To nie jest różnica stylu. To jest różnica architektury, z której styl dopiero wynika.

ZMYSŁ PERYFERYJNY: WIEDZIEĆ, KIEDY WYMIENIĆ KONTEKST

Żeby świadomie zarządzać życiem sesji, trzeba jedno: stale widzieć, co się w nich dzieje. Do tego służy Lambda Sensor - małe okno, które można trzymać w rogu ekranu i widzieć kątem oka, bez odrywania się od pracy: jak zapełnia się kontekst, ile zostało do kompaktacji, ile kosztowała ostatnia tura i jak zmienia się średnia. Zielony pasek postępu mówi wszystko, co trzeba: jest zapas - pracuj dalej; robi się ciasno - domknij wątek i wymień agenta na świeżego, zanim kompaktacja zrobi to za Ciebie po swojemu.

To drobiazg, który zmienia relację użytkownika z narzędziem: tam, gdzie inni dowiadują się o stanie kontekstu po fakcie - po dziwnym zachowaniu agenta albo po rachunku - tu widzisz go cały czas, zanim cokolwiek się wydarzy.

HIGIENA, KTÓRA SAMA SIĘ OPŁACA

Jest jeszcze jeden skutek tego modelu pracy, którego mało kto się spodziewa: niższe rachunki.

Każde narzędzie agentowe na rynku płaci ukryty podatek od zimnych startów. Pamięć podręczna promptów u dostawców modeli wygasa po przerwie - i wtedy pierwsza tura po powrocie kosztuje wielokrotność normalnej, bo cały kontekst jest przetwarzany od zera. Dzieje się to wszędzie: po każdej dłuższej pauzie, po każdym restarcie narzędzia. Tylko że nikt tego użytkownikowi nie pokazuje, więc ludzie widzą jedynie, że "czasem jest drożej", i nie wiedzą dlaczego.

Kreion działa na to z trzech stron naraz. Dba o cache promptów aktywniej niż jakiekolwiek inne narzędzie - sesje po prostu żyją w cieple dłużej, a kto chce, może włączyć pełną ochronę przed zimnymi startami jednym kliknięciem. Pokazuje wszystko na Lambda Sensor - każde trafienie w cache i każdy koszt jest widoczny, nie zakładany. A sama higiena sesji robi resztę: zimny start na małym, świeżym kontekście kosztuje ułamek zimnego startu na kontekście spuchniętym od tygodnia hodowli. Czyli nawet ten, kto nie włączy żadnej dodatkowej opcji, płaci mniej - samą dyscypliną pracy, którą ten model wymusza naturalnie.

DZIESIĘĆ MINUT NAUKI, INNY ZAWÓD

Cały ten model sprowadza się do kilku odruchów, których nauka zajmuje mniej czasu niż przeczytanie tego tekstu: logi do krótkich spraw, nowy wątek po przerwie - od New Session, projekt to szufladka z własnym agentem, konteksty zmieniaj jak rękawiczki, kątem oka patrz na Lambda Sensor, a wiedzę zostaw kartridżowi - on jej nie zgubi.

Kto przepracuje tak tydzień, przestaje rozumieć, jak mógł pracować inaczej: hodować jedną wieczną sesję ze strachu przed utratą wiedzy albo grzebać w cmentarzysku bezimiennych kontekstów w poszukiwaniu wczorajszej rozmowy. Świadome sterowanie życiem agentów - powoływanie ich i wymienianie bez sentymentu, bo ciągłość trzyma kartridż - to nie jest kolejna funkcja do opanowania. To jest inny zawód: przestajesz być opiekunem jednej kruchej rozmowy, a stajesz się szefem zespołu wymiennych specjalistów z dostępem do wspólnego, wiecznego archiwum.

I to jest dokładnie ten moment, w którym BTTF przestaje być "pamięcią między czatami", a staje się tym, czym jest naprawdę: fundamentem, który zmienia wszystko powyżej - łącznie z tym, czym w ogóle jest sesja.