

Inteligencja generatywna i inteligencja deklaratywno-poznawcza od 1920 roku w perspektywie antropologiczno-ewolucyjnej ssaków

Streszczenie wykonawcze

W niniejszym raporcie proponuję **operacyjne** (testowalne) rozróżnienie między *inteligencją generatywną* (IG) a *inteligencją deklaratywno-poznawczą* (IDP), a następnie śledzę ich **historyczny rozwój po 1920 r.** w psychologii, neuronauce, antropologii poznawczej i sztucznej inteligencji. Oś argumentu jest „mammalian-first”: najpierw mechanizmy uczenia i motywacji ssaków (zwłaszcza systemy dopaminergiczne), a potem warstwy kulturowe człowieka i analogie/rozbieżności wobec nowoczesnych systemów SaaS+AI. ¹

Najważniejsze tezy (w reżimie „twardej naukowości” – z zaznaczeniem, co jest danymi, a co wnioskowaniem): (a) IG i IDP nie są „dwoma organami” ani prostym podziałem na kreatywność vs pamięć, lecz **dwoma funkcjonalnymi pętlami**: *generuj→selekcjonuj→konsoliduj* (IG) oraz *koduj→przechowuj→odtwarzaj i kontroluj* (IDP). (b) U ssaków rdzeniem obu pętli jest uczenie przez wzmocnienie i sygnały wartości; w neurobiologii szczególnie istotne są dopaminergiczne sygnały błędu predykcji nagrody (RPE) oraz mechanizmy „chcenia” (wanting) odróżnialne od „lubienia” (liking). ²

Koncepcję „**piramidy dopaminowej**” oceniam jako **metaforę heurystyczną** popularną w obiegu internetowym, ale **nie jako termin ugruntowany w recenzowanej neuronauce**; jej największe ryzyko to „masłowizacja dopaminy” (sprowadzanie wielofunkcyjnego neuromodulatora do jednej osi przyjemności/motywacji). Jednocześnie można ją „uratować” metodologicznie: jako *schemat interwencji behawioralnych* (higiena snu, ruch, ograniczenie bodźców), ale bez przypisywania jej literalnej mechaniki dopaminowej. ³

Pokazuję, jak **symulacje Monte Carlo** (MC) i rodzina modeli uczenia przez wzmocnienie mogą formalizować procesy motywacji i decyzji (np. eksploracja–eksploatacja, preferencja nagród zmiennych, efekty „jackpotów”, wrażliwość na RPE). Podaję przykładowe projekty symulacji i oczekiwane wyniki, oraz prezentuję ilustracyjny wykres z prostego zadania bandyty (2-ramiennego). ⁴

W części porównawczej zestawiam strategię **Homo sapiens**, wilka i psa z logiką **SaaS+AI** (w tym mechanikami utrzymania użytkownika i „nagrodami” systemowymi). Wskazuję konwergencje (skalowalna kooperacja człowieka; koordynacja łowiecka wilka; międzygatunkowa kooperacja psa; „dystrybucja poznania” w systemach technicznych) oraz zasadnicze rozjazdy (brak ucieleśnienia i homeostazy w AI; inne „funkcje celu” w SaaS; podatność ssaczego układu nagrody na bodźce nadnormalne). ⁵

Ramowanie i definicje operacyjne

Założony „szkielet heurystyczny” użytkownika (bo nie został jawnie podany): w całym raporcie stosuję stałą strukturę, która zwykle wspiera pamięć i wnioskowanie: **Kotwica (definicja) → Mechanizm (jak działa) → Predykcje (co powinno być obserwowalne) → Test/kontrtest (co to falsyfikuje)**. Traktuję to jako minimalną wersję „struktur pamięciowo-heurystycznych” (chunkowanie, hierarchizacja, pętle sprzężeń zwrotnych, mapowanie analogii). Źródłem tej decyzji jest *brak specyfikacji heurystyk*, co jawnie odnotowuję także w sekcji „Założenia”. ⁶

Definicja inteligencji generatywnej

Inteligencja generatywna (IG): zdolność systemu poznawczego do **wytwarzania nowych (nie wprost odtworzonych) struktur**: hipotez, planów, narracji, narzędzi, rozwiązań lub zachowań, które są *jednocześnie* (1) **zróżnicowane** (variational), (2) **ukierunkowane** przez cele/ograniczenia środowiska i (3) **poddawane selekcji/ocenie** (wewnętrznej lub społecznej), po czym (4) **konsolidowane** w repertuarze jednostki lub kultury. W ujęciu psychologii kreatywności IG odpowiada za fazy „generowania” i „eksploracji/oceny” (modele typu Geneplore), a w neuronauce kreatywności – za dynamiczną współpracę sieci spontanicznej ideacji (DMN) i kontroli wykonawczej (ECN) w cyklu generuj–ocień. ⁷

Predykcje/operacjonalizacja IG: (a) lepsza wydajność w zadaniach dywergencyjnych (wiele rozwiązań), (b) większa elastyczność przełączania strategii, (c) większa zdolność do kontrfaktycznej symulacji (planowanie „co jeśli”), (d) w danych neuroobrazowych – wzrost współsprzężenia DMN–ECN w fazach generowania/oceny. ⁸

Definicja inteligencji deklaratoryjno-poznawczej

Inteligencja deklaratoryjno-poznawcza (IDP): zdolność do **jawnego kodowania, przechowywania i kontrolowanego odtwarzania** informacji o zdarzeniach i faktach (pamięć epizodyczna i semantyczna), oraz do ich **świadomej manipulacji** w rozumowaniu, komunikacji i planowaniu. IDP jest blisko spokrewniona z rozróżnieniem pamięci **deklaratywnej (jawnej)** i **niedeklaratywnej (niejawnej)** w neuropsychologii, oraz z tradycją modeli pamięci (magazynowej i roboczej). ⁹

Predykcje/operacjonalizacja IDP: (a) lepsze wyniki w zadaniach odtwarzania faktów i epizodów, (b) większa stabilność i „przenośność” wiedzy między kontekstami, (c) wyższa skuteczność kontroli poznawczej (utrzymanie celu, hamowanie reakcji), (d) większa skuteczność uczenia deklaratywnego w modelach pamięci (np. lepsza konsolidacja i odtwarzanie). ¹⁰

Relacja IG–IDP w ujęciu antropologicznym

W antropologii poznawczej i teoriach kultury kluczowe jest to, że **IDP nie jest tylko „w głowie”**, lecz bywa **rozszerzana przez artefakty** (pismo, narzędzia, mapy, procedury organizacyjne), a IG może działać „ponadjednostkowo” w zespołach jako poznanie rozproszone. W tym sensie „inteligencja” człowieka jest często **systemem człowiek-narzędzie-instytucja**, nie wyłącznie cechą mózgu. ¹¹

Geneza i rozwój pojęć od 1920 roku

Poniżej rekonstruuje rozwój idei, z których składają się współczesne konstrukty IG i IDP. W ramach „reżimu absolutnie naukowego” traktuję to jako **historię narzędzi pojęciowych**: pamięci, kreatywności/generatywności, uczenia przez wzmocnienie, oraz kultury jako mechanizmu ewolucyjnego. ¹²

timeline

title Oś czasu konceptów IG i IDP po 1920 r.

1926 : "The Language and Thought of the Child" (rozwój poznawczy i język)

1932 : "Remembering" (pamięć rekonstrukcyjna, schematy)

1934 : "Thinking and Speech" (społeczne źródła wyższych funkcji)

1949 : "The Organization of Behavior" (plastyczność i uczenie)

1950 : "Computing Machinery and Intelligence" + program badań kreatywności

1955-1957 : Narodziny AI jako programu badawczego + generatywność języka

1968-1974 : Modele pamięci magazynowej i roboczej

1972-2004 : Pamięć epizodyczna/semantyczna; systemy deklaratywne

1997-2011 : Dopamina/RPE; hipoteza mózgu społecznego; kultura jako

adaptacja

2011-2017 : SaaS (chmura) + architektury uczenia głębokiego

2022-2023 : RLHF w LLM + ramy zarządzania ryzykiem AI

Kamienie milowe „deklaratywności” (IDP) obejmują: klasyczne definicje pamięci jako kodowania–przechowywania–odtworzenia oraz rozróżnienie pamięci jawnej i niejawnej; w polskiej literaturze przeglądowej widoczne jest powiązanie podziałów pamięci z funkcją struktur mózgu oraz z modelami Atkinsona-Shiffrina i pamięci roboczej. ¹³

Kamienie milowe „generatywności” (IG) to: przejście od rozumienia inteligencji jako *wyłącznie* konwergentnego rozwiązywania zadań do rozumienia jej jako zdolności do **dywergentnej produkcji** (generowania wariantów), formalizowane w badaniach kreatywności; współcześnie wzmacniane przez neuronaukę kreatywności (współpraca sieci spontanicznych i kontrolnych) oraz przez modele predykcyjne mózgu (generatywne modele percepcji/inferencji). ¹⁴

Równolegle rozwijała się antropologiczna linia argumentu: człowiek ma szczególnie silną adaptację do kultury (uczenie społeczne, kumulacja wiedzy), a sama kultura może tworzyć nowe środowiska selekcyjne (koewolucja gen–kultura). To przesunęła ciężar „inteligencji” z jednostkowego mózgu ku mechanizmom transmisji i instytucjom. ¹⁵

W AI i informatyce od połowy XX wieku pojawia się drugi, techniczny sens „generatywności”: systemy zdolne do wytwarzania nowych ciągów symboli/tekstów na podstawie reguł lub modeli statystycznych. Ten sens w 2017 r. zyskał ultrasilny impuls dzięki architekturze transformera, a w 2022 r. – dzięki upowszechnieniu post-treningu i dopasowania do preferencji użytkowników (RLHF). ¹⁶

Ewolucyjne mechanizmy ssaków: socjalność, uczenie się, neurobiologia nagrody

Co znaczy „ramowanie ssacze”

U ssaków (w tym u człowieka) poznanie jest sprzęgnięte z **regulacją energii, ryzyka i relacji społecznych**. Dlatego najbardziej „realistyczny” fundament pod IG i IDP to nie abstrakcyjne IQ, lecz **pętla: bodziec→decyzja→działanie→konsekwencja→aktualizacja**. W neurobiologii te pętla są osadzone w układach neuromodulatorowych (m.in. dopaminergicznym) oraz w sieciach pamięci i kontroli. ¹⁷

Dopamina jako sygnał uczenia i hierarchie nagród

W recenzowanej literaturze dopamina jest ściśle wiązana z funkcjami **uczenia o wartości**, w szczególności z błędem predykcji nagrody (RPE): różnicą między tym, co oczekiwane, a tym, co otrzymane. Klasyczna praca w *Science* z 1997 r. (często cytowana jako neuronowy substrat przewidywania i nagrody) formalnie łączy sygnały dopaminowe z uczeniem typu reinforcement learning.

18

Równolegle ważna jest korekta potocznego mitu „dopamina = przyjemność”. Modele „incentive salience” rozdziela **wanting** (motywacyjne „chcenie”, napędzane m.in. przez szlaki mezolimbiczne) od **liking** (hedoniczne „lubienie”), które nie musi zależeć bezpośrednio od dopaminy. To rozróżnienie jest kluczowe, gdy analizujemy mechaniki aplikacji, gamifikacji czy „detoksów dopaminowych”. 19

W polskim piśmiennictwie przeglądowym znajdziemy dodatkowo jednoznaczne uporządkowanie: receptory dopaminowe dzielą się na rodziny D1-like i D2-like o przeciwnych efektach na sygnalizację komórkową; oraz podkreślenie, że system dopaminergiczny współdziała z innymi układami (serotonina, noradrenalina, glutaminian, GABA, endokannabinoidy, opioidy), co utrudnia proste „jednowymiarowe” schematy. 20

Hierarchie nagród w rzeczywistości są wielowymiarowe: pokarm, bezpieczeństwo, status, więzi, seks, eksploracja. Z punktu widzenia antropologii ewolucyjnej szczególnie ważne są nagrody społeczne, bo u gatunków stadnych dostęp do zasobów jest często zapośredniczony przez **pozycję w strukturze grupy**. Prace na naczelnych pokazują, że status społeczny może współwystępować ze zmianami dostępności receptorów D2; to sugeruje mechanizm, w którym „hierarchia społeczna” jest równocześnie „hierarchią neurochemiczną” wpływającą na podatność na wzmocnienia. 21

Socjalność i „mózg społeczny” jako tło IG/IDP

Hipoteza mózgu społecznego wiąże rozmiar/architekturę kory (zwłaszcza neokorteksu) z obciążeniem obliczeniowym utrzymywania relacji społecznych. Choć szczegóły (w tym interpretacje liczby Dunbara) bywają dyskutowane, rdzeń hipotezy jest ważny dla tego raportu: **inteligencja ssacza jest w dużej mierze inteligencją relacyjną** – wymaga pamięci o innych (IDP) i generowania zachowań w strategii społecznej (IG). 22

W ludzkiej linii ewolucyjnej część tej presji selekcyjnej została „przeskalowana” przez kulturę: uczenie społeczne i instytucje umożliwiły kumulację rozwiązań (co jest osiowo generatywne) oraz stabilne repozytoria wiedzy (co jest osiowo deklaratywne). Prace o „kulturowej kognicji” i „koewolucji gen-kultura” traktują te procesy jako realne mechanizmy ewolucyjne, a nie metaforę. 23

Piramida dopaminowa i modelowanie Monte Carlo motywacji

„Piramida dopaminowa” jako pojęcie: rekonstrukcja i ocena naukowa

W obiegu internetowym „dopamine pyramid / piramida dopaminowa” zwykle oznacza hierarchię zachowań, które mają „podnosić dopaminę” (fundament: sen/ruch/światło; wyżej: muzyka, hobby; najwyżej: media społecznościowe itp.). Empirycznie to pojęcie **nie funkcjonuje jako standardowy termin neuronaukowy** – raczej jako infografika lifestyle’owa. (To stwierdzenie opieram na charakterze dominujących wyników oraz na tym, że debaty naukowe koncentrują się raczej na funkcjach dopaminy i uczeniu, nie na „piramidach”). 24

Najbliższy naukowy „kuzyn” tej metafory to nie piramida, lecz: (a) teoria RPE (dopamina jako sygnał uczenia), (b) teoria incentive salience (wanting vs liking), (c) mechanizmy harmonogramów wzmocnień (zmienne nagrody → wysokie tempo reakcji), oraz (d) badania nad wpływem przerw/ograniczeń bodźców na dobrostan, które bywają błędnie sprzedawane jako „detoks dopaminowy”. W polskim materiale popularyzatorskim podkreśla się, że „detoks dopaminowy” opiera się często na pseudonaukowych interpretacjach, nawet jeśli ograniczenie bodźców może mieć korzyści psychologiczne. ²⁵

Werdykt metodologiczny:

- Jako **model neurobiologiczny**: „piramida dopaminowa” jest zbyt słaba (brak definicji zmiennych, brak mechanizmu, brak falsyfikowalnych predykcji; ryzyko pomylenia dopaminy z przyjemnością). ²⁶
- Jako **heurystyka behawioralna**: może być użyteczna, jeśli zostanie przeformatowana na „piramidę interwencji” (od podstawowych regulatorów: sen/ruch/rytmy dobowe → po świadome zarządzanie bodźcami), bez twierdzeń o „resetowaniu dopaminy”. ²⁷

Monte Carlo jako język formalny dla motywacji

Metody Monte Carlo (MC) w klasycznym sensie to **losowe próbkowanie** do przybliżania wielkości trudnych analitycznie; kanoniczna publikacja z 1949 r. opisuje MC jako statystyczne podejście do problemów równań i zjawisk fizycznych. ²⁸

W uczeniu przez wzmocnienie MC ma szczególne znaczenie, bo pozwala estymować wartość działań przez **średnią z obserwowanych zwrotów (returns)** bez jawnego modelu środowiska; w klasycznym podręczniku RL MC jest jedną z trzech fundamentalnych rodzin metod obok programowania dynamicznego i uczenia TD. ²⁹

W planowaniu sekwencyjnym (gdzie „motywacja” = wybór trajektorii) MC rozwinęło się w kierunku Monte-Carlo Tree Search oraz algorytmów typu UCT, które łączą próbkowanie trajektorii z zasadą eksploracja–eksploatacja. To daje formalny analog do ssaczej strategii: *eksploruj nowe zasoby vs eksploatuj znane*. ³⁰

Przykładowe projekty symulacji i oczekiwane wyniki

Projekt A: „Bandyta nagród zmiennych” (model podatności na wzmocnienia nieregularne)

- *Stan*: brak (2-ramienny bandyta).
- *Akcje*: wybór A (stabilna nagroda) lub B (rzadkie „jackpoty”).
- *Uczenie*: aktualizacja wartości Q przez błąd ($rew - Q$).
- *Parametry interpretowalne biologicznie*: szybkość uczenia α (wrażliwość na RPE), oraz temperatura/ β w softmax (skłonność do eksploatacji).
- *Oczekiwane wyniki*: przy pewnych ustawieniach agent zaczyna preferować ramię o wysokiej wariancji (B), mimo że jego średnia wartość bywa podobna lub niższa – analogicznie do tego, jak zmienne wzmocnienia mogą podtrzymywać zachowanie. To spójne z literaturą o harmonogramach wzmocnień i z rozdzieleniem wanting/liking. ³¹

Wykres z symulacji

Projekt B: „Hierarchia potrzeb jako wieloskładnikowa funkcja nagrody” (pseudopiramida → model testowalny)

- *Stan*: wektor zasobów (energia, bezpieczeństwo, afiliacja, status, nowość).
- *Akcje*: zachowania „krótkie” (natychmiastowa gratyfikacja) i „długie” (inwestycja w stabilność).

- *Nagroda*: suma ważona, z wagami zależnymi od deficytu (homeostaza) oraz od sygnałów społecznych.
- *MC*: symulacja długich trajektorii i rozkładów wyników (np. ryzyko „pułapki dopaminowej”: przeoptymalizowanie krótkich nagród).
- *Oczekiwane wyniki*: pojawienie się faz, w których system „przestawia wagi” (np. stres zwiększa preferencję szybkich wzmocnień), co jest zgodne z obserwacją, że układ dopaminergiczny działa w silnych interakcjach ze stresem i innymi układami neuroprzebiegowymi. ³²

Projekt C: „Uczenie statusu społecznego” (hierarchie nagród społ.)

- *Stan*: relacje dyadyczne w grupie + pamięć epizodyczna interakcji.
- *Nagroda*: dostęp do zasobów i redukcja ryzyka konfliktu.
- *Mechanizm*: uczenie polityki zachowań dominacyjnych/ugodowych.
- *Oczekiwane wyniki*: stabilizacja hierarchii; wrażliwość parametrów „dopaminowych” na podatność na ryzykowne wzmocnienia. Jako empiryczne tło: dane o związku statusu z dostępnością receptorów D2 u naczelnych. ³³

Porównanie strategii: Homo sapiens, wilk, pies i SaaS+AI

Punkt wyjścia: SaaS+AI jako system selekcyjny

SaaS (Software as a Service) w definicji instytucjonalnej to możliwość korzystania z aplikacji dostawcy działających w infrastrukturze chmurowej, dostępnych przez interfejs klienta (np. przeglądarkę). Ten model jest ważny, bo wprowadza „ekologię” użytkownika: ciągłe aktualizacje, metryki retencji, eksperymenty A/B i pętle wzmacniania zachowań. ³⁴

W warstwie AI współczesne generatywne systemy (np. LLM) są wzmacniane przez post-trening z preferencjami człowieka (RLHF) oraz przez ramy zarządzania ryzykiem (np. AI RMF). To wprowadza formalny odpowiednik „selekcji” – nie biologicznej, lecz produktowo-rynkowej i regulacyjnej. ³⁵

Tabela porównawcza kluczowych atrybutów

Wymiar	Homo sapiens	Wilk (dziki)	Pies (udomowiony)	SaaS+AI (system produktowy)
Rdzeń poznawczy	Silna kumulacja kulturowa; sprzężenie generowania idei z kontrolą i pamięcią jawną	Poznanie ukierunkowane na koordynację łowiecką i rodzinny system stada	Wzmocnione kompetencje międzygatunkowe (komunikacja z człowiekiem); zmiany temperamentu	Generacja treści/planów statystycznych + optymalizacja celu (metryki); brak homeostazy biologicznej
Struktura społeczna	Wielopoziomowe sieci; instytucje; koalicje; prestiż i normy	Zwykle rodzina: rodzice prowadzą; praca dzielona; mit „walki o alfę” bywa nadmierny	Silna więź z człowiekiem; elastyczne układy społ. zależne od środowiska	„Społeczność użytkowników” + infrastruktura; relacja klient–dostawca; pętle retencji

Wymiar	Homo sapiens	Wilk (dziki)	Pies (udomowiony)	SaaS+AI (system produktowy)
Uczenie	Silne uczenie społeczne; język i artefakty jako pamięć zewnętrzną	Uczenie społeczne w stadzie + doświadczenie środowiskowe	Selekcja na współpracę z człowiekiem; specyficzne odczytywanie wskazówek ludzkich	Uczenie maszynowe: pre-trening, fine-tuning, RLHF; testy online/A-B
System nagrody / „wartość”	Układ nagrody ssaczy (RPE, wanting/liking) + nagrody społeczne i instytucjonalne	Nagrody biologiczne i społeczne w ramach stada i polowania	Nagrody biologiczne + silne wzmocnienia społeczne od człowieka	Funkcja celu (np. zysk/retencja) + sygnały zwrotne od użytkownika; ryzyko projektowania bodźców nadnormalnych
Typowe „ryzyka niedopasowania”	Nadstymulacja bodźcami nowoczesnymi; konflikty celów krótkie vs długie	Konflikty zasobowe; presja środowiska; mityzacja hierarchii w captiva	Konflikty z wilkami i człowiekiem; błędne odczytanie potrzeb i socjalizacji	Optymalizacja metryk zamiast dobrostanu; halucynacje/ „konfabulacje”; zewnętrżność odpowiedzialności

Źródłowe kotwice dla treści tabeli: rodzinna struktura stada wilków i krytyka uproszczonej narracji „alpha” (Mech), specyfika psiej kognicji społecznej względem wilków, genomika domestykacji i adaptacje, definicja SaaS oraz ramy i techniki współczesnej AI (NIST/OECD/RLHF). ³⁶

Konwergencje, rozbieżności i implikacje praktyczne

Konwergencje (realne podobieństwa funkcjonalne):

- 1) *Eksploracja-eksploatacja*: wilk eksploruje teren, ale eksploatuje znane strategie łowieckie; człowiek eksploruje idee i instytucje; SaaS+AI eksploruje warianty interfejsu i modeli przez eksperymenty oraz uczenie, a LLM „próbkuje” alternatywne kontynuacje. ³⁷
- 2) *Dystrybucja poznania*: człowiek i jego organizacje wyprowadzają poznanie poza jednostkę (narzędzia, zespoły), co ma analog w architekturach chmurowych i w praktykach MLOps; to jest zgodne z teoriami poznania rozproszonego i „rozszerzonego umysłu”. ³⁸
- 3) *Socjalne sygnały jako wzmocnienia*: zarówno u ssaków stadnych, jak i u ludzi, nagrody społeczne modulują zachowanie; w środowiskach cyfrowych analogicznie działają „lajki”, komentarze i inne sygnały, które można modelować w języku uczenia nagrodowego. ³⁹

Rozbieżności (miejsca, gdzie analogie pękają):

- 1) *Brak homeostazy w AI*: ssaczy układ nagrody jest sprzęgnięty z metabolizmem i kosztami biologicznymi; AI (w szczególności w SaaS) optymalizuje funkcję celu bez wbudowanej „fizjologii” dobrostanu użytkownika, chyba że zostanie to explicite dodane w wymaganiach/standardach. ⁴⁰
- 2) *Wanting bez liking w projektowaniu produktów*: w systemach cyfrowych można łatwo wzmacniać „chcenie” (pętla bodziec→sprawdź→zmienna nagroda), nawet jeśli doświadczenie nie jest realnie satysfakcjonujące; to jest dokładnie obszar, gdzie potoczne mity o dopaminie mylą trop. ⁴¹

3) *Kooperacja wilków vs psów*: dane porównawcze sugerują, że wilki mogą być bardzo tolerancyjne i kooperatywne z pobratymcami, podczas gdy u psów selekcja domestykacyjna mogła przesunąć kooperację szczególnie w stronę człowieka (zależnie od zadania). To oznacza, że „pies jako model człowieka” jest użyteczny tylko warunkowo. ⁴²

Implikacje praktyczne (dla projektowania SaaS+AI w zgodzie z realnością ssaczą):

- Projektowanie interfejsów powinno uwzględniać, że **wzmocnienia zmienne** i bodźce społeczne mogą tworzyć silne pętle nawykowe; etycznie oznacza to potrzebę ograniczeń, przejrzystości i testowania skutków ubocznych (dobrostan, kompulsywność), co jest spójne z podejściem „risk management” do AI.

⁴³
- W generatywnej AI należy rozdzielać: (a) zdolność generowania (IG-podobną) od (b) stabilnej „wiedzy deklaratywnej” i weryfikacji (IDP-podobnej). Praktycznie oznacza to architektury z retrieval-augmented generation, audytem źródeł i mechanizmami oceny prawdziwości, bo sama generacja nie gwarantuje trafności. (To jest wniosek integrujący znane ograniczenia modeli generatywnych i praktyki post-treningu). ⁴⁴

- W pracy z człowiekiem jako „systemem ssaczym” warto traktować „piramidę dopaminową” jak checklistę higieny bodźców, ale decyzje wdrożeniowe opierać na modelach uczenia/nagród oraz na danych o zachowaniach użytkownika (np. analizy RL zachowań w social media). ⁴⁵

Założenia i pytania otwarte

Założenia przyjęte jawnie (bo brakowało danych wejściowych)

1) **Heurystyki użytkownika** nie zostały podane, więc zastosowałem minimalny, deklaratywnie opisany „szkielet pamięciowo-heurystyczny” (kotwica→mechanizm→predykcje→test) i konsekwentnie go użyłem. (To nie jest twierdzenie o użytkowniku, tylko o strukturze raportu). ⁴⁶

2) Terminów „inteligencja generatywna” i „inteligencja deklaratywno-poznawcza” używam jako **konstruktów syntetycznych** zbudowanych na fundamencie uznanych pól (psychologia kreatywności, pamięć deklaratywna, antropologia poznawcza, neuronauka dopaminy). Zakładam, że użytkownik oczekuje takiej syntezy, bo prosi o „kreatywne, lecz rygorystyczne” ujęcie. ⁴⁷

3) W części SaaS+AI zakładam, że interesuje nas **typowy** model produktowy (retencja, iteracyjne eksperymenty) i **typowe** techniki post-treningu modeli (np. RLHF jako przykład), a nie specyficzna firma czy jeden produkt. ⁴⁸

4) Symulacje MC traktuję jako **narzędzie ilustracyjne** do pokazania mechanizmów (nie jako model kliniczny ani empirycznie skalibrowany opis człowieka). ⁴⁹

Pytania otwarte, które są najbardziej „nośne” badawczo

Jak najlepiej ilościowo rozdzielać IG i IDP u ludzi w sposób, który jest jednocześnie porównywalny międzykulturowo (antropologia) i neurobiologicznie ugruntowany (sieci DMN/ECN, pamięć deklaratywna)? Obszar kreatywności ma narzędzia (zadania dywergencyjne, sieci funkcjonalne), ale ich transfer kulturowy jest nieoczywisty. ⁵⁰

Czy „wanting/liking” można operacyjnie mapować na specyficzne metryki w produktach SaaS (np. częstotliwość sprawdzania vs satysfakcja deklarowana) i włączyć to do ram zarządzania ryzykiem AI jako standardowe ryzyko behawioralne? To łączy neuronaukę nagrody, analizę zachowań w mediach społecznościowych i standardy AI. ⁵¹

W jakim stopniu różnice wilk-pies w kooperacji są funkcją ekologii (warunki zadania, środowisko wychowania) vs skutkiem selekcji domestykacyjnej, i kiedy analogia „pies jako model współpracy

człowieka” jest uprawniona? Najlepsze dane wskazują na zależność od kontekstu i partnera (człowiek vs pobratymiec), co ogranicza proste analogie. ⁴²

Czy da się zbudować „piramidę” nie dopaminy, lecz **wielowarstwowej funkcji wartości** (homeostaza, afiliacja, status, ciekawość) i testować ją eksperymentalnie (np. przez parametry RL, czasy reakcji, wybory ryzykowne)? To byłaby naukowa wersja tego, co popkultura próbuje opisać metaforą „piramidy dopaminowej”. ⁵²

¹ ¹⁷ <https://www.ncbi.nlm.nih.gov/books/NBK232964/>

<https://www.ncbi.nlm.nih.gov/books/NBK232964/>

² ¹⁸ A Neural Substrate of Prediction and Reward

https://www.gatsby.ucl.ac.uk/~dayan/papers/sdm97.pdf?utm_source=chatgpt.com

³ ²⁵ ²⁷ ⁴⁵ <https://zdrowie.pap.pl/psyche/detoks-dopaminowy-prawdy-i-mity>

<https://zdrowie.pap.pl/psyche/detoks-dopaminowy-prawdy-i-mity>

⁴ ²⁸ ⁴⁹ <https://people.bordeaux.inria.fr/pierre.delmoral/MetropolisUlam49.pdf>

<https://people.bordeaux.inria.fr/pierre.delmoral/MetropolisUlam49.pdf>

⁵ ³⁴ ⁴⁸ Software as a Service (SaaS) - Glossary

https://csrc.nist.gov/glossary/term/software_as_a_service?utm_source=chatgpt.com

⁶ ¹⁰ ¹² ¹³ ⁴⁶ <https://ppn.ipin.edu.pl/pl/zeszyty/archiwum/2006-tom-15-zeszyt-2/wspolczesne-modele-pamieci-w-aspekcie-neurobiologicznym-i-klinicznym/download>

<https://ppn.ipin.edu.pl/pl/zeszyty/archiwum/2006-tom-15-zeszyt-2/wspolczesne-modele-pamieci-w-aspekcie-neurobiologicznym-i-klinicznym/download>

⁷ ¹⁴ ⁴⁷ <https://ecologylab.net/research/publications/WardSmithFinke.pdf>

<https://ecologylab.net/research/publications/WardSmithFinke.pdf>

⁸ ⁵⁰ <https://pmc.ncbi.nlm.nih.gov/articles/PMC4410786/>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC4410786/>

⁹ Episodic and Semantic Memory

https://alicekim.ca/EMSM72.pdf?utm_source=chatgpt.com

¹¹ ³⁸ https://uberty.org/wp-content/uploads/2015/07/Edwin_Hutchins_Cognition_in_the_Wild.pdf

https://uberty.org/wp-content/uploads/2015/07/Edwin_Hutchins_Cognition_in_the_Wild.pdf

¹⁵ ²³ https://www.eva.mpg.de/documents/Cambridge/Tomasello_Understanding_BehBrainSci_2005_1555292.pdf

https://www.eva.mpg.de/documents/Cambridge/Tomasello_Understanding_BehBrainSci_2005_1555292.pdf

¹⁶ <https://arxiv.org/abs/1706.03762>

<https://arxiv.org/abs/1706.03762>

¹⁹ ⁴¹ ⁵¹ <https://pmc.ncbi.nlm.nih.gov/articles/PMC5171207/>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC5171207/>

²⁰ ³² ⁵² <https://www.psychiatriapolska.pl/pdf-80500-79518?filename=Funkcjonowanie+ukladu.pdf>

<https://www.psychiatriapolska.pl/pdf-80500-79518?filename=Funkcjonowanie+ukladu.pdf>

²¹ ³³ ³⁹ <https://pubmed.ncbi.nlm.nih.gov/11802171/>

<https://pubmed.ncbi.nlm.nih.gov/11802171/>

22 <https://cognitionandculture.net/wp-content/uploads/Evolutionary-Anthropology-1998-Dunbar-The-social-brain-hypothesis.pdf>

<https://cognitionandculture.net/wp-content/uploads/Evolutionary-Anthropology-1998-Dunbar-The-social-brain-hypothesis.pdf>

24 <https://www.instagram.com/p/DTs1csBD3zz/>

<https://www.instagram.com/p/DTs1csBD3zz/>

26 https://www.lecerveau.ca/flash/capsules/articles_pdf/dopamine.pdf

https://www.lecerveau.ca/flash/capsules/articles_pdf/dopamine.pdf

29 <https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf>

<https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf>

30 37 <https://ggp.stanford.edu/readings/uct.pdf>

<https://ggp.stanford.edu/readings/uct.pdf>

31 43 <https://pmc.ncbi.nlm.nih.gov/articles/PMC1473025/>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC1473025/>

35 44 <https://cdn.openai.com/papers/>

[Training_language_models_to_follow_instructions_with_human_feedback.pdf](https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf)

https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf

36 https://www.wolf.org/wp-content/uploads/2013/09/267alphastatus_english.pdf

https://www.wolf.org/wp-content/uploads/2013/09/267alphastatus_english.pdf

40 <https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>

<https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>

42 <https://www.nature.com/articles/s41598-019-40468-y>

<https://www.nature.com/articles/s41598-019-40468-y>