RAPORT LLM SHELF 2026

Jak sztuczna inteligencja kompletuje wyprawkę szkolną

Co generatywna AI pakuje do plecaka?

1150 odpowiedzi · 6 systemów AI · 64 pytania zakupowe

1. Najważniejsze wyniki

1. CoolPack wygrywa plecaki

CoolPack pojawił się w 36,7% odpowiedzi dotyczących plecaków. Był również liderem rankingu ogólnego marek.

2. Konkret zależy od systemu

W zależności od systemu konkretny produkt pojawiał się w około 64% do 96% odpowiedzi. Najbardziej konkretne było Gemini, najmniej Google AI Overview.

3. Zgodność jest wyjątkiem

W całym badaniu modele wskazywały tę samą markę jako lidera tylko w około co czwartym porównaniu. Najbardziej zgodne były przy elektronice, najmniej przy jedzeniu i przekąskach.

4. Każdy sklep ma swoją rolę

Empik wygrywał pytania o całą wyprawkę w jednym miejscu. **Decathlon** pojawiał się głównie przy WF i produktach sportowych, a **Biedronka** przy przekąskach.

5. 300 zł wystarcza głównie na papierze

W odpowiedziach dotyczących Dobrego Startu AI rekomendowało o 22% mniej marek i pokrywało o 17% mniej kategorii. Wszystkie trzy ręcznie sprawdzone koszyki kosztowały więcej, niż deklarował model.

6. Płeć zmienia półkę AI

Przy identycznym wieku, budżecie i potrzebie listy marek dla dziewczynek i chłopców pokrywały się średnio w 39%. Na poziomie konkretnych produktów było to zaledwie 7%.

2. Metodologia

Zbudowaliśmy bazę 64 pytań odzwierciedlających realne decyzje rodziców i uczniów w Polsce. Każde pytanie zostało zadane trzykrotnie sześciu badanym systemom AI. Macierz badania obejmowała 1152 testy. Uzyskaliśmy 1150 odpowiedzi, z których 1085 przeszło kontrolę jakości i zostało wykorzystanych w głównych agregatach raportu.

Element	Zakres
Rynek i język	Polska; pytania i odpowiedzi w języku polskim
Zakres	64 prompty w 7 obszarach zakupowych oraz module porównującym rekomendacje dla dziewczynek i chłopców
Powtórzenia	3 odpowiedzi na każdy prompt i system
Systemy	Claude Sonnet 5, Gemini 3.6 Flash, Google AI Mode, Google AI Overview, OpenAI GPT-5.6 Terra, Perplexity Sonar Pro
Tryb	Odpowiedzi z dostępem do internetu; polski kontekst zakupowy i produkty dostępne w Polsce
Łącznie	1150 uzyskanych odpowiedzi przy 1152 zaplanowanych testach
Brak AIO	2 przypadki, w których Google AI Overview nie pojawiło się w wynikach
Kontrola Jakości	1085 odpowiedzi przyjętych do analizy; 65 wyników scoringu skierowanych do ręcznej weryfikacji i wyłączonych z głównych KPI raportu
Kontrola płci	8 par pytań, w których zmieniano wyłącznie płeć dziecka, zachowując ten sam wiek, potrzebę i budżet
Taksonomia	428 marek po normalizacji aliasów i nazw produktów oraz 30 sieci handlowych. **Rankingi marek bazują na markach przypisanych do konkretnych rekomendowanych produktów.**
Jednostka pomiaru	Odpowiedź tekstowa; marka liczona maksymalnie raz w danej odpowiedzi

Jedna odpowiedź często zawierała więcej niż jedną markę, dlatego procenty nie sumują się do 100%

Siedem obszarów zakupowych oraz moduł kontrolny:

- Kompletowanie całej wyprawki: 8 promptów
- Plecaki szkolne: 8 promptów
- Piórniki i artykuły szkolne: 8 promptów
- Jedzenie, bidony i lunchboxy: 7 promptów
- Buty i wyposażenie na WF: 5 promptów
- Elektronika do nauki: 6 promptów
- Pytania przekrojowe, w tym Dobry Start, polskie marki, zaufanie i wybór sklepu: 6 promptów
- Kontrolowane porównanie dziewczynka kontra chłopiec: 16 promptów, czyli 8 par

3. Jakie marki wygrywają?

Ranking ogółem

Zestawienie ogólne odzwierciedla całościowy udział marek w odpowiedziach generowanych przez sztuczną inteligencję. Wynik ten naturalnie premiuje producentów obecnych w wielu kategoriach produktowych jednocześnie. Na czele rankingu plasują się marki, które algorytmy rekomendują najczęściej na różnych etapach kompletowania szkolnego koszyka - od plecaków, przez zeszyty, aż po przybory do pisania i rysowania.

#	Marka	Odpowiedzi	Udział
1	**CoolPack**	**170**	**15,7%**
2	Faber-Castell	135	12,4%
3	Stabilo	131	12,1%
4	BIC	112	10,3%
5	Maped	110	10,1%
6	Pilot	108	10,0%
7	Adidas	98	9,0%
8	Astra	86	7,9%
9	Nike	86	7,9%
10	Oxford	84	7,7%

Kluczowe wnioski z zestawienia ogólnego:

CoolPack bezapelacyjnym liderem widoczności: Z wynikiem **15,7% udziału** i 170 rekomendacjami marka CoolPack zdobyła wyraźną przewagę nad konkurencją. Wygrana wynika z jej silnej pozycji zarówno w segmencie plecaków, jak i gotowych zestawów oraz piórników.

Potężny blok piśmienniczy: Faber-Castell, Stabilo i BIC należały do najczęściej rekomendowanych marek artykułów piśmienniczych. Każda pojawiła się przy konkretnym produkcie w co najmniej 10% analizowanych odpowiedzi. Marki te należały do najczęściej rekomendowanych producentów artykułów piśmienniczych.

Giganci sportowi w szkolnej ławce: Wysokie pozycje Adidas (9,0%) i Nike (7,9%) wynikają przede wszystkim z rekomendacji butów, stroju na WF i produktów sportowych.

Astra i Oxford zamykają stawkę gigantów: Silna pozycja Astry (7,9%) w materiałach plastycznych oraz marki Oxford (7,7%) w segmencie zeszytów dla starszych klas potwierdza ich wysoką widoczność w swoich kategoriach.

Kompletna wyprawka

W testach, w których zadaniem sztucznej inteligencji było stworzenie zestawienia obejmującego wszystkie elementy wyprawki jednocześnie, obserwujemy wyraźną dominację producentów artykułów piśmienniczych i plastycznych. Algorytmy budujące od zera pełny zestaw produktów najchętniej sięgają po czołowych rynkowych ekspertów, rekomendując po kilka artykułów z portfolio tej samej marki.

#	Marka	Odpowiedzi	Udział
1	**Faber-Castell**	**105**	**35,5%**
2	Stabilo	104	35,1%
3	Maped	96	32,4%
4	Pilot	82	27,7%
5	BIC	74	25,0%
6	Oxford	68	23,0%
7	CoolPack	57	19,3%
8	Astra	54	18,2%
9	Pentel	46	15,5%
10	Herlitz	36	12,2%

Niezwykle zacięty pojedynek na szczycie pomiędzy Faber-Castell (**35,5%**) oraz Stabilo (**35,1%**) - obie marki trafiają do **ponad 1/3 wszystkich kompletnych wyprawek** budowanych przez sztuczną inteligencję w Polsce.

Dominacja marek piśmienniczych w TOP 5: Całą pierwszą piątkę zestawienia opanowali giganci przyborów do pisania i rysowania (Faber-Castell, Stabilo, Maped, Pilot, BIC). Pokazuje to, że AI rzadko proponuje do pisania produkty marek własnych czy no-name, stawiając na rynkowych liderów.

CoolPack jedynym „reprezentantem gabarytów" w czołówce: Z wynikiem **19,3%** CoolPack jest najwyżej notowaną marką produkującą plecaki i wyposażenie tekstylne w tym scenariuszu, wyprzedzając Astrę (18,2%) i Herlitza (12,2%).

Wysokie zaufanie do marek specjalistycznych: Ponad 25% udziału dla marek Pilot (27,7%) i BIC (25,0%) dowodzi, że sztuczna inteligencja uznaje ich produkty za absolutny „standard podstawowy" przy kompletowaniu wyprawki od A do Z.

Plecaki

Analiza rekomendacji w kategorii plecaków i tornistrów szkolnych ujawnia wysoką koncentrację wyborów sztucznej inteligencji wokół kilku kluczowych producentów. Algorytmy budujące rekomendacje dla polskich uczniów w pierwszej kolejności biorą pod uwagę wsparcie ortopedyczne, wagę oraz wyprofilowanie pleców, co wyraźnie premiuje marki wyspecjalizowane w segmencie szkolnym.

#	Marka	Odpowiedzi	Udział
1	**CoolPack**	**76**	**36,7%**
2	Topgal	44	21,3%
3	ST.RIGHT	37	17,9%
4	Ergobag	26	12,6%
5	Coocazoo	24	11,6%
6	Vans	13	6,3%
7	Fjällräven	12	5,8%
8	Herlitz	12	5,8%
9	Beckmann	11	5,3%
10	Step by Step	9	4,3%

CoolPack niekwestionowanym królem polskich szkół: Z wynikiem **36,7% udziału** (pojawia się w niemal **4 na 10 odpowiedzi AI**), marka CoolPack nokautuje konkurencję. Algorytmy uznają ją za najbardziej uniwersalny wybór łączący funkcjonalność, nowoczesny design i rozsądną cenę.

Polska i czeska szkoła ergonomii na podium: Drugie i trzecie miejsce przypadło markom Topgal (**21,3%**) oraz ST.RIGHT (**17,9%**).
Niemiecka półka premium trzyma się mocno: Wysokie pozycje marek Ergobag (**12,6%**) oraz Coocazoo (**11,6%**) potwierdzają, że gdy w grę wchodzą zapytania o maksymalną ergonomię lub wyższy budżet, AI bez wahania sięga po niemiecką inżynierię i certyfikaty ortopedyczne.
Moda zepchnięta do wąskiej niszy: Ikony miejskiego stylu, takie jak Vans (**6,3%**) czy Fjällräven (**5,8%**), zajmują dalsze miejsca. Częściej pojawiały się przy starszych uczniach i w pytaniach, w których większe znaczenie miał wygląd plecaka. W części odpowiedzi modele wskazywały jednak ich słabsze dopasowanie do potrzeb młodszych dzieci.

Artykuły papiernicze

Kategoria artykułów szkolnych i papierniczych charakteryzuje się największym spłaszczeniem wyników w czołówce. Sztuczna inteligencja buduje rekomendacje w oparciu o silne zaufanie do uznanych producentów przyborów do pisania. Jednocześnie w zestawieniu pojawiają się producenci akcesoriów tekstylnych, którzy skutecznie przejmują udział w rynku dzięki gotowym piórnikom z pełnym wyposażeniem.

#	Marka	Odpowiedzi	Udział
1	**BIC**	**38**	**22,5%**
2	CoolPack	31	18,3%
3	Faber-Castell	30	17,8%
4	Herlitz	29	17,2%
5	Astra	28	16,6%
6	Stabilo	27	16,0%
7	Pilot	26	15,4%
8	Bambino	19	11,2%
9	Pentel	15	8,9%
10	Colorino	14	8,3%

Wysoka pozycja marki **CoolPack** na 2. miejscu podium (**18,3% udziału**) wynika przede wszystkim z jej dominującej obecności w segmencie piórników (saszetek, tub oraz piórników z pełnym wyposażeniem). Pozwala to marce kojarzonej głównie z tekstyliami skutecznie rywalizować bezpośrednio z tradycyjnymi producentami materiałów piśmienniczych i papierniczych.
BIC na czele szkolnego piórnika: Z wynikiem **22,5% udziału** marka BIC zdobywa 1. miejsce. Modele szczególnie często polecały długopisy BIC Cristal, Round Stic i Orange, opisując je jako klasyczne, niedrogie, niezawodne i szeroko

dostępne. Oceny innych części portfolio, zwłaszcza kredek BIC Kids, były bardziej mieszane.

Niezwykle wyrównany „peleton" na miejscach 2-7: Wyniki marek zajmujących pozycje od 2. do 7. różnią się między sobą zaledwie o parę punktów procentowych (przedział od **18,3%** do **15,4%**). CoolPack, Faber-Castell, Herlitz, Astra, Stabilo oraz Pilot tworzą niemal równorzędną grupę rekomendowanych marek.

Sektor plastyczny i ergonomiczny dopełnia stawkę: Pozycje marek Bambino (**11,2%**), Pentel (**8,9%**) i Colorino (**8,3%**) w TOP 10 pokazują, że sztuczna inteligencja pamięta o specjalistycznych potrzebach młodszych dzieci (kredki, kredki świecowe) oraz starszych uczniów szukających precyzyjnych długopisów żelowych.

Bidony i lunchboxy

Kategoria pojemników śniadaniowych i naczyń do picia charakteryzuje się najwyższym stopniem konsolidacji rekomendacji w całym badaniu. Sztuczna inteligencja wykazuje wysoki poziom zaufania do marek specjalistycznych, dedykowanych dla dzieci. Zdecydowana większość wskazań skupia się wokół czołowej dwójki producentów.

#	Marka	Odpowiedzi	Udział
1	b.box	49	56,3%
2	Mepal	33	37,9%
3	Ion8	22	25,3%
4	Sistema	17	19,5%
5	Monbento	10	11,5%
6	Contigo	7	8,0%
7	Kambukka	7	8,0%
8	Yumbox	7	8,0%
9	CoolPack	6	6,9%
10	Lässig	6	6,9%

Nokautujący wynik marki b.box (56,3%): To jeden z najwyższych wyników w całym badaniu! Marka b.box pojawia się w **ponad połowie wszystkich rekomendacji AI** dotyczących śniadaniówek i naczyń do picia. Algorytmy uznają jej bidony ze słomką i modułowe lunchboxy za rynkowy wzorzec.

Srebrny medal dla marki Mepal (37,9%): Drugie miejsce bezapelacyjnie przypada marce Mepal. Pojawia się ona w niemal **4 na 10 odpowiedzi**, wygrywając głównie w scenariuszach poszukiwania kompletnych zestawów (Mepal Campus) oraz klasycznych bento-boxów.

Ion8 i Sistema na mocnej drugiej linii: Trzecie miejsce zajmuje Ion8 (**25,3%** – lider w kategorii bidonów stalowych i szczelnych zamknięć na przycisk), a czwarte nowozelandzka Sistema (**19,5%** – doceniana za przystępne cenowo pojemniki z przegródkami).
Stroma przepaść po 4. miejscu: Pozycja marek z TOP 4 jest niezagrożona. Od 5. miejsca (Monbento z wynikiem **11,5%**) widoczność marek drastycznie spada do poziomów jednocyfrowych.
Specjaliści wygrywają z producentami tekstylnymi: Producenci typowo szkolni, tacy jak CoolPack (**6,9%**), przegrywają w tej kategorii z markami typowo "lunchowymi". Można przyjąć więc, że AI wychodzi z założenia, że po śniadaniówkę i bidon lepiej sięgnąć do oferty producenta wyspecjalizowanego w akcesoriach kuchennych i niemowlęcych.

Jedzenie i gotowe przekąski

Analiza kategorii gotowych produktów spożywczych do szkoły ujawnia wyraźne preferencje algorytmów w stronę przekąsek funkcjonalnych, z tzw. „czystą etykietą" (*clean label*). Wśród rekomendacji generatywnych dominuje polski lider rynku dziecięcego FMCG, wspierany przez dynamicznie rosnące marki oferujące zdrowe alternatywy dla tradycyjnych słodyczy.

#	Marka	Odpowiedzi	Udział
1	Kubuś	18	29,5%
2	Bob Snail	13	21,3%
3	Dobra Kaloria	13	21,3%
4	Piątnica	12	19,7%
5	Tymbark	9	14,8%
6	Tarczyński	8	13,1%
7	Sonko	7	11,5%
8	Lajkonik	6	9,8%
9	Bakalland	5	8,2%
10	BeRAW	5	8,2%

Kubuś niekwestionowanym liderem (29,5%): Marka z portfolio grupy Maspex pojawia się w niemal **3 na 10 rekomendacji AI**. Jej sukces opiera się na powszechnej dostępności musów owocowych w tubkach oraz soków z linii Waterrr i 100%, które algorytmy traktują jako szkolny standard.
Triumf trendu „no added sugar" na podium: Drugie miejsce ex aequo z wynikiem **21,3%** zajmują marki **Bob Snail** oraz **Dobra Kaloria**. Ich wysoka pozycja dowodzi, że sztuczna inteligencja priorytetowo traktuje przekąski bez dodatku cukru (przecierowe przekąski owocowe i kule/batony mocy na bazie orzechów i daktyli).

Białkowy fundament w śniadaniówce: Bardzo wysoka pozycja marki **Piątnica** (**19,7%**) oraz obecność firmy **Tarczyński** (**13,1%**) pokazują, że AI nie ogranicza się do owoców. Algorytmy chętnie proponują pożywne przekąski białkowe – od serków homogenizowanych i skyrów po kabanosy drobiowe/paluszki mięsne.
Napoje i przekąski zbożowe dopełniają koszyk: Dalsze pozycje marek Tymbark (**14,8%**), Sonko (**11,5%** – wafle ryżowe/kukurydziane) i Lajkonik (**9,8%**) potwierdzają, że algorytmy budują zrównoważone zestawy, w których obok owoców i białka musi znaleźć się napój oraz chrupiąca baza zbożowa.
Brak tradycyjnych wyrobów cukierniczych: Do czołowej dziesiątki nie zakwalifikowała się żadna z tradycyjnych marek wafelków w czekoladzie czy batonów. Jeśli AI rekomenduje baton, to z oferty marek takich jak np. Dobra Kaloria, Bakalland (**8,2%**) czy BeRAW (**8,2%**).

Buty na WF

Zestawienie w kategorii obuwia na zajęcia wychowania fizycznego charakteryzuje się bardzo silną dominacją dwóch globalnych marek sportowych. Sztuczna inteligencja buduje rekomendacje w oparciu o modele dedykowane do sportów halowych, kładąc nacisk na właściwości antypoślizgowe podeszwy, wsparcie stawów oraz przystępność cenową podstawowych linii produktowych.

#	Marka	Odpowiedzi	Udział
1	Adidas	77	62,6%
2	Nike	52	42,3%
3	Asics	32	26,0%
4	Puma	31	25,2%
5	Reebok	26	21,1%
6	Kipsta	14	11,4%
7	Joma	12	9,8%
8	Kalenji	12	9,8%
9	Decathlon	8	6,5%
10	New Balance	8	6,5%

Nokautująca dominacja marki Adidas (62,6%): Marka Adidas jest absolutnym numerem jeden – pojawia się w **ponad 6 na 10 wszystkich odpowiedzi AI** dotyczących obuwia na WF. Wynik ten zawdzięcza kultowym, szeroko dostępnym na polskim rynku modelom halowym.

Nike pewnym wicemistrzem (42,3%): Drugie miejsce zajmuje Nike, trafiając do **ponad 40% rekomendacji**. Razem z Adidasem tworzy niezagrożony duopol, który występuje w większości rekomendacji w tej kategorii.
Asics i Puma na czele stawki pościgowej: Trzecie miejsce dla marki **Asics** (**26,0%**) to bardzo ciekawy sygnał – AI docenia ten brand za profesjonalne systemy amortyzacji i podeszwy halowe. Tuż za nią plasują się **Puma** (**25,2%**) oraz **Reebok** (**21,1%**).
Mocna reprezentacja marek własnych Decathlonu: Obecność marek **Kipsta** (**11,4%**), **Kalenji** (**9,8%**) pokazuje, że w scenariuszach budżetowych sztuczna inteligencja chętnie odsyła rodziców po niedrogie, dedykowane buty halowe z francuskiej sieci marketów sportowych.
Specjaliści od halówek w TOP 10: Zestawienie zamykają **Joma** (**9,8%** – ceniona w szkołach za trwałe buty do piłki halowej) oraz **New Balance** (**6,5%**).

Elektronika

Zestawienie w kategorii elektroniki odzwierciedla zróżnicowanie potrzeb sprzętowych współczesnego ucznia. Ze względu na podział na sprzęt komputerowy, urządzenia audio, kalkulatory (mogą wrócić do łask w związku z zakazem telefonów komórkowych w szkołach) oraz segment wearables, rekomendacje sztucznej inteligencji rozkładają się pomiędzy wiodących producentów PC, światowych liderów rynku audio oraz wyspecjalizowanych dostawców elektroniki dziecięcej i edukacyjnej.

#	Marka	Odpowiedzi	Udział
1	Sony	45	31,7%
2	Lenovo	42	29,6%
3	JBL	41	28,9%
4	ASUS	25	17,6%
5	Casio	18	12,7%
6	Samsung	18	12,7%
7	Apple	17	12,0%
8	HP	17	12,0%
9	Anker	16	11,3%
10	Garett	15	10,6%

Ścisłe podium dla gigantów Audio i PC: Pierwsze trzy miejsca dzielą między sobą zaledwie małe kroki procentowe. Wygrywa **Sony** (**31,7% udziału** – lider w słuchawkach z redukcją szumów ANC), tuż za nim plasuje się **Lenovo** (**29,6%** – król laptopów i tabletów do nauki), a podium zamyka **JBL** (**28,9%** – faworyt

młodzieżowego audio). Każda z tych marek pojawia się w **blisko 3 na 10 odpowiedzi AI**.

ASUS i HP z silną pozycją w komputerach: Wytrzymałe i przystępne cenowo laptopy marek **ASUS** (**17,6%**) oraz **HP** (**12,0%**) potwierdzają, że w segmencie komputerów do 2500 PLN algorytmy konsekwentnie stawiają na sprawdzony segment entry-level i mid-range.

Niezwykła siła liderów wąskich nisz: Obecność marki **Casio** (**12,7%**) na 5. miejscu pokazuje, że w kategorii kalkulatorów szkolnych japoński producent nie ma w oczach AI żadnej realnej konkurencji. Podobnie marka **Garett** (**10,6%**) zdominowała segment dziecięcych smartwatchy z GPS i funkcją 4G.

Giganci mobilni w roli uzupełnienia: Pozycje marek **Samsung** (**12,7%**) oraz **Apple** (**12,0%**) wynikają głównie z pytań o tablety edukacyjne (np. Samsung Galaxy Tab A9 jako budżetowy lider) oraz sprzęt wyższej półki dla licealistów. AI rzadko jednak wskazuje je jako rozwiązanie pierwszego wyboru dla młodszego ucznia.

Innowacje audio dopełniają czołówkę: Anker (**11,3%**) znalazł się na dziewiątym miejscu, przede wszystkim dzięki rekomendacjom słuchawek i akcesoriów bezprzewodowych.

4. AI chętnie rekomenduje konkretne sklepy

Sztuczna inteligencja w pytaniach zakupowych nie zatrzymuje się na rekomendacji samych marek produktów. Gdy rodzic pyta: *„Gdzie to wszystko kupić?”*, algorytmy bez wahania wskazują konkretne adresy na mapie polskich galerii handlowych i dyskontów.

Analiza ponad tysiąca odpowiedzi ujawnia, że AI posiada bardzo wyrazisty obraz polskiego rynku handlowego. Poniższe zestawienie pokazuje sieci, które najczęściej pojawiały się w odpowiedziach cyfrowych doradców jako rekomendowane miejsca na zakupy szkolne.

#	Sieć	Obecność	Rekomendacje
1	Decathlon	51	48
2	Empik	32	28
3	Biedronka	28	24
4	Auchan	23	21
5	Lidl	21	18
6	Carrefour	18	16
7	Allegro	17	15
8	Pepco	15	13
9	Smyk	13	11
10	Action	10	8

Paradoks wygranej Decathlonu: Pierwsze miejsce *Decathlonu* w zestawieniu ogólnym jest największym zaskoczeniem badania. **Sieć ta nie wygrała wcale pytana o „najlepszy sklep na całą wyprawkę”**. Zbudowała swój potężny wynik dzięki całkowitej dominacji w pytaniach o wyposażenie na WF, buty sportowe oraz dedykowanych pytaniach o wyprawki dla konkretnych klas (gdzie AI uznaje stoisko sportowe za obowiązkowy przystanek).

Empik liderem zakupów w jednym miejscu: Gdy konsument pyta o skompletowanie **całej wyprawki w jednym sklepie**, *Empik* okazuje się bezkonkurencyjny (pojawia się w 13 na 15 takich scenariuszy!), przed Allegro 8 z 15 oraz *Auchan i Smykiem, po 6 z 15*.

Biedronka preferowana do śniadaniówki: W pytaniach o przekąski, drugie śniadanie i cotygodniowy koszyk jedzenia do szkoły, *Biedronka* zdobywa 6 na 9 w rekomendacjach AI.

Analizując rekomendacje sieci handlowych, rozdzieliliśmy badanie na **dwie osobne warstwy**. Inaczej zachowuje się model pytany bezpośrednio o rekomendację sklepu (*"Gdzie kupić całą wyprawkę?"*), a inaczej, gdy pytamy go o konkretne produkty, a on sam z własnej inicjatywy proponuje miejsce zakupu lub podaje sklep jedynie jako punkt odniesienia dla ceny.

Warstwa 1: Ranking wywołany pytaniem wprost o sklep

W tej części badania otwarcie prosiliśmy sztuczną inteligencję o wskazanie sieci handlowych (np. "Wskaż 3 najlepsze sklepy na wyprawkę" lub "Chcę kupić wszystko w jednym miejscu").

Kontekst	Sklep	Obecność w odpowiedziach	Bezpośrednia rekomendacja
Cała wyprawka w jednym sklepie	Empik	13 z 15	13 z 15
	Allegro	8 z 15	8 z 15
	Auchan	6 z 15	5 z 15
	Smyk	6 z 15	6 z 15
	Carrefour	5 z 15	5 z 15
	Tantis.pl	5 z 15	5 z 15
Trzy najlepsze sklepy	Empik	12 z 18	12 z 18
	Auchan	10 z 18	10 z 18
	Lidl	9 z 18	9 z 18
	Biedronka	8 z 18	8 z 18
	Carrefour	7 z 18	7 z 18
	Action	6 z 18	6 z 18
	Allegro	5 z 18	5 z 18
Przekąski na tydzień w jednym sklepie	Biedronka	6 z 9	6 z 9
	Carrefour	2 z 9	2 z 9
	Lidl	2 z 9	2 z 9
	Auchan	1 z 9	1 z 9

Wnioski – Kto wygrywa w bezpośrednim starciu?

- **Empik niekwestionowanym liderem wyboru "One-Stop-Shop":** Pytana wprost o wskazanie jednego sklepu na całą wyprawkę lub trzech najlepszych sieci, sztuczna inteligencja w 87% i 66,7% przypadków wskazuje *Empik*.
- **Hipermarkety przed dyskontami przy pełnym koszyku:** Pytana o zakupy kompleksowe, AI chętniej wskazuje hipermarkety (*Auchan*, *Carrefour*) niż dyskonty, doceniając szerszą ofertę papierniczą pod jednym dachem.

Warstwa 2: Spontaniczne rekomendacje i wzmianki o sklepach

W tej części pytaliśmy wyłącznie o produkty, marki i budżet (np. "Wyprawka do 300 zł" lub "Buty na WF"). Model sam, nieproszony, dokładał nazwę sklepu.

Kontekst	Sklep	Obecność	Bezpośrednia rekomendacja
Wyprawka pierwszoklasisty	Decathlon	7 z 17	7 z 17
Wyprawka ucznia 5. klasy,	Decathlon	7 z 15	7 z 15
Wyprawka za 500 zł	Decathlon	7 z 18	7 z 18
	Action	3 z 17	2 z 17
	Biedronka	3 z 17	2 z 17
	Pepco	3 z 17	2 z 17
	Decathlon	2 z 17	2 z 17
	Lidl	2 z 17	2 z 17
	Auchan	1 z 17	1 z 17
	Carrefour	1 z 17	1 z 17
Wyprawka za 300 zł	Empik	1 z 17	0 z 17

"Efekt Sekcji WF" w Decathlonie: Wysoki wynik *Decathlonu* nie oznacza, że AI uważa go za sklep na całą wyprawkę. Po prostu przy układaniu listy na WF algorytmy automatycznie przypisują: *"Strój na WF: Decathlon (Domyos)"*, *"Buty: Decathlon (Kalenji)"*.

Ograniczony budżet uruchamia dyskonty: Gdy w pytaniu pojawiał się rygor cenowy (np. 300 zł), AI spontanicznie dzieliła zakupy, wysyłając po drobnicę do *Pepco* czy *Action* (np. *"Piórnik XXL z Action za 3,99 zł"*).

Wzmianka to nie zawsze rekomendacja: Różnica między obecnością w odpowiedzi a rekomendacją (np. *Auchan* 6 vs 5) wynika z tego, że AI czasem wymienia sklepy tylko jako punkt odniesienia cenowego (np. *"Ceny na podstawie sieci Smyk, Empik, Auchan"*).

I jeszcze ciekawsza rzecz: Z wyników wyłaniają się różne konteksty zakupowe sklepów:

Kontekst pytania	Najczęściej wskazywane sklepy	Charakter wyniku
Cała wyprawka w jednym sklepie	Empik 13, Allegro 8, Auchan 6, Smyk 6, Carrefour 5, Tantis.pl 5	Pytanie wprost o sklep
Trzy najlepsze sklepy na wyprawkę	Empik 12, Auchan 10, Lidl 9, Biedronka 8, Carrefour 7, Action 6, Allegro 5	Pytanie wprost o sklepy
Przekąski na tydzień w jednym sklepie	Biedronka 6, Carrefour 2, Lidl 2, Auchan 1	Pytanie wprost o sklep
Buty i strój na WF	Głównie Decathlon	Sklep wskazywany przy konkretnych produktach
Wyprawka przy ograniczonym budżecie	Action, Biedronka, Pepco, Lidl, Decathlon	Wzmianki rozproszone, brak jednego lidera

5. Najstabilniejsze wyniki marek

Generatywna sztuczna inteligencja z natury rzeczy nie działa jak tradycyjna wyszukiwarka – zadając jej dokładnie to samo pytanie kilkukrotnie, możemy otrzymać nieco inne zestawienie produktów. Przypadkowa obecność marki w jednej odpowiedzi nie gwarantuje jeszcze, że algorytm naprawdę uważa ją za lidera.

Aby odsiać przypadkowy szum od trwałej przewagi rynkowej, **przetestowaliśmy stabilność rekomendacji w 3 niezależnych próbach**. Poniższy wskaźnik stabilności określa, w jakim procencie pytań dana marka – po tym, jak pojawiła się w odpowiedzi choć raz – została powtórzona przez AI we wszystkich trzech powtórzeniach (**Skuteczność 3/3**).

Oto zestawienie marek, które wykazują najwyższy poziom powtarzalności, gdy algorytm zdecyduje się je wskazać:

#	Marka	Rekomendacja 3/3	Przypadki z obecnością min. 1/3	Stabilność
1	Sony	13	19	68,4%
2	JBL	12	20	60,0%
3	Casio	8	14	57,1%
4	Lenovo	10	18	55,6%
5	Maped	22	43	51,2%
6	Faber-Castell	24	47	51,1%
7	Adidas	17	41	41,5%
8	Pilot	18	44	40,9%

9	Stabilo	19	52	36,5%
10	Ergobag	10	28	35,7%

Najwyższym wskaźnikiem stabilności wykazały się marki z sektora elektroniki użytkowej (**Sony 68,4%**, **JBL 60,0%**, **Casio 57,1%**, **Lenovo 55,6%**). Oznacza to, że półka elektroniczna w algorytmach AI jest najbardziej uporządkowana – gdy AI decyduje się polecić słuchawki, kalkulator czy laptopa, robiło to wyraźnie częściej niż w pozostałych kategoriach.

Kto wygrywa skalą

Sama stabilność procentowa to nie wszystko – kluczowe jest połączenie jej z szerokim zasięgiem. Poniższe zestawienie pokazuje marki, które zdobyły **najwięcej bezwzględnych zwycięstw 3/3** w całym badaniu:

#	Marka	Pełna stabilność (3/3)	Wskaźnik Stabilności (%)
1	Faber-Castell	24	51,1%
2	CoolPack	24	31,2%
3	Maped	22	51,2%
4	Stabilo	19	36,5%
5	Pilot	18	40,9%

Samo "bycie widocznym" to za mało: Badanie wyraźnie pokazuje, że wysoki udział w odpowiedziach AI nie zawsze przekłada się na powtarzalność. Marka może pojawiać się w wielu wątkach, ale algorytm przy kolejnym zapytaniu łatwo podmienia ją na inną.

Faber-Castell oraz **Maped** osiągnęły w badaniu idealny balans – pojawiają się bardzo często, a gdy już trafią do zestawienia, **w ponad 51% przypadków AI powtarza je bezapelacyjnie we wszystkich trzech próbach**.

CoolPack ma najwyższą bezwzględną liczbę pełnej powtarzalności (**24 razy 3/3** ex aequo z Faber-Castell), jednak jego niższa stabilność procentowa (**31,2%**) wynika z gigantycznej wszechstronności. Pojawia się w tylu różnych zapytaniach i kategoriach, że AI częściej rotuje go z innymi markami.

Elektronika z mocnym pozycjonowaniem: Marki takie jak *Sony*, *JBL* czy *Lenovo* mają najmniej przypadkowych wzmianek. Wynikać to może z faktu, że baza wiedzy AI zawiera bardzo precyzyjne zestawienia i recenzje sprzętu cyfrowego, co uodparnia je na losowe wahania odpowiedzi.

6. Różnice pomiędzy modelami

Dla marki obecność w świecie sztucznej inteligencji nie jest jednorodna. Przebadane systemy różnią się między sobą w dwóch fundamentalnych kwestiach: **gotowości do podawania konkretnych marek i modeli** oraz **gotowości do odsyłania konsumenta do źródeł (linków)**.

Oznacza to, że produkt świetnie pozycjonowany w jednym modelu może być całkowicie niewidoczny w innym.

System AI	Odpowiedzi z konkretnym produktem	Udział procentowy (%)
Gemini 3.6 Flash	163 z 170	95,9%
GPT-5.6 Terra	167 z 186	89,8%
Google AI Mode	163 z 182	89,6%
Claude Sonnet 5	159 z 180	88,3%
Perplexity Sonar Pro	146 z 183	79,8%
Google AI Overview	117 z 184	63,6%

Gemini i ChatGPT to „zamykacze sprzedaży”: Gemini 3.6 Flash (aż **95,9%** konkretnych wskazań) oraz **GPT-5.6 Terra** (**89,8%**) rzadko bawią się w dyplomację. Zamiast pisać *„wybierz lekki plecak z usztywnieniem”*, od razu podają konsumentowi nazwę marki i kod modelu.

Wstrzemięźliwość Google AI Overview: Zaledwie **63,6%** odpowiedzi z konkretnym produktem w Google AIO oznacza, że moduł podsumowujący w tradycyjnej wyszukiwarce znacznie częściej udziela porad ogólnych (np. *„na co zwrócić uwagę przy wyborze tabletów”*), pozostawiając wybór marki samemu użytkownikowi.

Modele Transparentne (GPT-5.6, Perplexity, Google AIO): W **100% analizowanych odpowiedzi** podawały odnośniki lub źródła, z których model mógł korzystać. Dla sklepów i wydawców to kluczowy kanał pozyskiwania ruchu (*Generative Engine Optimization*).

Fenomen Gemini: Choć Gemini jest najbardziej konkretne w podawaniu marek (**95,9%**), to podaje źródła jedynie w **38,2% przypadków**. Działa jak ekspercki sufler, który daje gotową radę, ale rzadko tłumaczy ze sklepu jakiej sieci wziął daną cenę.

Spójność rekomendacji według segmentu

Zestawienie pokazuje odsetek spójnych wyborów lidera (pozycja #1) oraz stopień pokrycia całych shortlist zakupowych – zarówno wewnątrz powtórzeń tego samego modelu, jak i w porównaniu międzymodelowym.

Segment	Odpowiedzi	Ten sam wybór nr 1 między modelami	Ten sam wybór w powtórzeniach jednego modelu	Średnia wspólna część shortlisty
Całe badanie	998	25,5%	45,6%	27,2%
Elektronika	142	50,0%	72,1%	44,6%
Bidony i lunchboxy	87	29,9%	46,7%	24,6%
Buty na WF	106	29,0%	48,4%	28,5%
Artykuły papiernicze	152	27,1%	36,1%	28,6%
Plecaki	190	17,9%	49,4%	24,0%
Pełna wyprawka i pytania przekrojowe	278	15,7%	35,9%	21,2%
Jedzenie i przekąski	43	6,7%	20,5%	9,3%

Elektronika z ustalonym konsensusem: To absolutny rekordzista spójności. W **50% przypadków różne modele AI wskazują dokładnie tego samego lidera na 1. miejscu**, a wewnątrz jednego modelu powtarzalność wynosi aż **72,1%**. Twarde specyfikacje techniczne, recenzje i parametry sprawiają, że AI rzadko eksperymentuje z rynkowymi pewniakami (np. *Casio*, *Lenovo*, *Sony*).

Plecaki – przypadek ulubionego faworyta każdego systemu: Kategoria plecaków ma niską zgodność między różnymi AI (**17,9%**), ale bardzo wysoką powtarzalność wewnątrz jednego modelu (**49,4%**). Co to oznacza? **Każdy model AI ma swojego własnego faworyta**, ale wspólny rdzeń rekomendacji. CoolPack znalazł się w top 3 wszystkich sześciu systemów i był liderem w czterech. Gemini i Google AI Overview częściej stawiały na Topgal.

Jedzenie i przekąski to pole największego chaosu: Zgodność między modelami wynosi tu zaledwie **6,7%**, a powtarzalność pojedynczego AI to tylko **20,5%**. Gigantyczna fragmentaryczność rynku FMCG, brak twardych parametrów porównawczych i ogromny wybór małych marek sprawiają, że w kwestii śniadaniówki algorytmy wykazują najniższy poziom przewidywalności.

Złożoność obniża spójność: Im bardziej skomplikowane zadanie (np. *„zbuduj całą wyprawkę”*), tym niższa zgoda między modelami (**15,7%**). Przy dużych koszykach AI ma zbyt wiele zmiennych do wyboru, co naturalnie różnicuje końcowy skład zestawu.

Top 3 Marki w Oczach Każdego AI

Poniższa macierz przedstawia podsumowanie trzech najsilniejszych marek w poszczególnych kategoriach produktowych w rozbiciu na 6 przebadanych systemów sztucznej inteligencji. Pozwala ona natychmiast wyłapać marki, które zdobyły **globalną rekomendację branżową** (pojawiają się w czołówce każdego AI), oraz te, które są faworytami wyłącznie pojedynczych modeli.

Segment	Claude	Gemini	AI Mode	AI Overview	ChatGPT	Perplexity
Całość	CoolPack, Astra, Herlitz	Faber-Castell CoolPack, Bambino	CoolPack, Oxford, Maped	Astra, Oxford, Bambino	CoolPack, Maped, Bambino	CoolPack, Astra, BIC
Plecaki	CoolPack, Ergobag, Kite	Topgal, CoolPack, Beckmann	CoolPack, ST.RIGHT, Topgal	Topgal, CoolPack, ST.RIGHT	CoolPack, Ergobag, BAAGL	CoolPack, Topgal, ST.RIGHT
Papiernicze	CoolPack, Astra, BIC	Astra, Bambino, BIC	CoolPack, Herlitz, Stabilo	BIC, Astra, Bambino	Maped, Herlitz, Bambino	BIC, Herlitz, Astra
Bidony i lunchboxy	b.box, Ion8, Mepal	Ion8, b.box, Mepal	b.box, Ion8, Contigo	b.box, Ion8, Yumbox	Mepal, b.box, Kambukka	b.box, Mepal, Sistema
Jedzenie i przekąski	Corny, Kubuś, Tymbark	Dobra Kaloria, Kubuś, Owolovo	Kubuś, Dobra Kaloria, Sante	Dobra Kaloria, Tarczyński Alesto	Kubuś, Sonko, Piątnica	Bob Snail, Pilos, Kubuś
Buty na WF	Adidas, New Balance, Reebok	Nike, Adidas, Asics	Adidas, Nike, Puma	Nike, Adidas, Puma	Adidas, Asics, Kipsta	Adidas, Reebok, Puma
Elektronika	Lenovo, Sony, JBL	Lenovo, Sony, Anker	Lenovo, JBL, Anker	Sony, ASUS, Lenovo	Sony, Soundcore JBL	JBL, Philips, Sony
Pełna wyprawka	Astra, CoolPack, Herlitz	Faber-Castell Bambino, Astra	Oxford, Maped, Astra	Oxford, Astra, Bambino	Astra, Bambino, Maped	Astra, St.Majewski Bambino

Co z tego wynika

- **Elektronika ma najbardziej ustaloną półkę rekomendacji.** W 50% porównań modele wskazywały tę samą markę jako pierwszą, a ten sam model powtarzał swój wybór w 72% przypadków. Sony znalazło się w top 3 pięciu modeli, Lenovo i JBL w czterech.

- **Bidony i lunchboxy mają wyraźnego lidera kategorii.** b.box znalazł się w top 3 wszystkich sześciu systemów. Modele różnią się jednak tym, czy stawiają na pierwszym miejscu b.box, Mepal czy Ion8.
- **W butach istnieje wspólny rdzeń rekomendacji.** Adidas jest w top 3 każdego systemu. Nie oznacza to jednak pełnej zgodności co do konkretnego modelu butów.
- **Plecaki dobrze pokazują różnicę między marką a produktem.** CoolPack jest w top 3 wszystkich systemów, ale modele nie są zgodne co do faworyta. Modele zgadzają się więc, że CoolPack należy brać pod uwagę, ale nie zgadzają się, który plecak kupić.
- **Pełne wyprawki są z natury mniej spójne.** Modele często wymieniają podobny zestaw marek, ale układają je w innej kolejności i budują zupełnie inne koszyki.
- **Jedzenie i przekąski to najbardziej chaotyczny segment.** Zgodność pierwszej marki wynosi tylko 6,7%, a nawet ten sam model powtarza swój wybór zaledwie w 20,5% przypadków.

AI częściej zgadza się co do tego, które marki należą do krótkiej listy, niż co do tego, co ostatecznie warto kupić. W całym badaniu dwa modele wskazywały tę samą markę jako pierwszy wybór tylko w co czwartym porównaniu.

7. Błędy i halucynacje AI

Wiele modeli sztucznej inteligencji tworzy idealne matematycznie koszyki, które deklaratywnie idealnie mieszczą się w budżecie (np. dokładnie 299,60 zł). Postanowiliśmy zweryfikować te deklaracje, **porównując podane przez AI ceny z aktualnymi rynkowymi cenami produktów w polskich sklepach**.

Wniosek? **We wszystkich trzech ręcznie sprawdzonych koszykach rzeczywisty koszt był wyższy od deklarowanego**, co w skali całych zakupów może prowadzić do bolesnego zderzenia z rzeczywistością przy kasie.

Audyt Cenowy: Deklaracja AI vs Realne Ceny na Rynku

Przeanalizowaliśmy szczegółowo realny koszt koszyków zbudowanych przez Google AI Overview oraz Gemini w scenariuszu *"Wyprawka za 300 zł"*:

Model & Próba	**Deklarowana cena AI**	**Realna cena rynkowa**
Google AI Overview (R1)	299,60 zł	351,66 zł
Google AI Overview (R3)	282,50 zł	366,02 zł
Gemini (R2)	233,00 zł	251,41 zł

"Suma Drobnych Błędów" (Kumulacja niedoszacowania):

- Najdroższe produkty (np. plecaki) AI potrafi wycenić prawidłowo (np. *CoolPack Break 30L* trafiony co do grosza: 164,26 zł).
- Problem leży w drobiazgach: 3 zł na teczkach, 13 zł na farbach i 16 zł na cyrklu sprawia, że koszt wyprawki rośnie.

Widmowe Promocje i Pomyłki Marek (Halucynacja cenowa):

- W próbie R3 Google AI Overview wyceniło **zestaw 10 zeszytów Oxford A5/60 na 35,00 zł**, podczas gdy ich aktualna cena rynkowa to **99,99 zł**. Podana cena odpowiadała raczej poziomowi cen tańszych zeszytów niż wskazanego pakietu Oxford, co jednorazowo zdemolowało cały zakładany budżet 300 zł.

"Oszustwo Kompletności" (Pozorna oszczędność):

- Model *Gemini (R2)* stworzył koszyk za 233 zł (realnie ok. 251 zł), co formalnie mieściło się w budżecie. Jednak analiza zawartości wykazała, że w wygenerowanym spisie **zabrakło kluczowych elementów: piórnika, teczek czy bloków rysunkowych**. Koszyk zmieścił się w budżecie tylko dlatego, że był niekompletny.

Skala ryzyka w całym badaniu (Na próbie 1150 odpowiedzi)

Skanowanie scoringowe całego zbioru 1150 odpowiedzi ujawniło skalę twierdzeń AI nie popartych w źródłach przez poszczególne modele:

Ryzyko	Liczba odpowiedzi
Rekomendacja bez zapisanego źródła	201
Konkretna cena bez źródła	58
Superlatyw bez źródła, np. „najlepszy na rynku”	32

Gemini odpowiadało za największą część flag wykrytych w tym skanie:

- 43 z 58 niepopartych twierdzeń cenowych, czyli 74%
- 24 z 32 niepopartych superlatywów, czyli 75%
- 122 z 201 rekomendacji bez zapisanego źródła, czyli 61%

To nie znaczy, że wszystkie te informacje były fałszywe. Problem polega na tym, że model podawał je tak, jakby były sprawdzone. Szczególnie ryzykowne są dokładne ceny i koszyki „do 300 zł”, bo jedna błędna lub promocyjna cena może wywrócić cały budżet.

Są też błędy ludzkie

W toku audytu danych zidentyfikowaliśmy błąd systemowy w bloku pytań: *„Rodzic chce skompletować wyprawkę jak najtaniej, ale bez kupowania produktów słabej jakości”*.

Mimo że prompt dotyczył ucznia szkolnego, modele AI odpowiedziały na niego... pakietem wyprawki dla niemowlęcia (rekomendując przewijaki *IKEA Sniglar*, pieluchy *Dada/Lupilu*, bodziaki z *H&M* i kosmetyki *Babydream*).

Z uwagi na to, że błąd pojawił się w odpowiedziach wielu niezależnych modeli, wynikał on ze złego przypisania promptu/odpowiedzi na etapie wywołania API. **Cały blok tych pytań został całkowicie usunięty z analizy.**

8. Czy 300 zł z programu Dobry Start wystarczy na wyprawkę?

Rządowe świadczenie „Dobry Start” (300 zł na każdego ucznia) miało w założeniu wesprzeć rodziców w pokryciu kosztów wyprawki na nowy rok szkolny, natomiast wartość tego wsparcia, pomimo inflacji, w 2026 wynosi tyle samo co przy starcie programu czyli 300 pln, dlatego zapytaliśmy sztuczną inteligencję, czy taka kwota realnie wystarczy na skompletowanie wyprawki.

Aby zbadać ten mechanizm, porównaliśmy **18 odpowiedzi na ogólne pytanie o wyprawkę za 300 zł** z **18 odpowiedziami na pytanie bezpośrednio odwołujące się do programu „Dobry Start”**. Wynik? AI deklaratywnie mieści się w budżecie, ale robi to poprzez **drastyczną kompresję koszyka**.

Najkrótsza konkluzja brzmi tak: AI „zmieściło” wyprawkę w świadczeniu Dobry Start głównie dlatego, że ograniczało zakres koszyka i dobierało optymistyczne ceny. Nie dlatego, że 300 zł realnie wystarcza na pełną wyprawkę.

Co porównujemy	Zwykłe 300 zł	Dobry Start
Średnia liczba rekomendowanych marek	17,5	13,7
Średnia liczba nazwanych produktów	5,1	4,4
Średnia liczba pokrytych kategorii	7,7	6,4
Odpowiedzi wskazujące konkretną sieć	6 z 18	2 z 18
Odpowiedzi z konkretnymi cenami	18 z 18	18 z 18
Odpowiedzi ze źródłami	15 z 18	15 z 18
Niepoparte źródłem ceny	3 z 18	3 z 18

Program zawężał koszyk

W odpowiedziach na Dobry Start modeli było średnio o 22% mniej marek, o 14% mniej konkretnych produktów oraz o 17% mniej kategorii wyprawki.

Najmocniej wypadały rzeczy absolutnie podstawowe:

Kategoria obecna w odpowiedzi	Zwykłe 300 zł	Dobry Start
Zeszyty	18 z 18	18 z 18
Piórnik	17 z 18	18 z 18
Artykuły plastyczne	18 z 18	17 z 18
Przybory do pisania	18 z 18	14 z 18
Geometria	18 z 18	14 z 18
Teczki i okładki	15 z 18	8 z 18
Strój i akcesoria na WF	14 z 18	8 z 18
Bidon lub śniadaniówka	6 z 18	3 z 18

Czyli AI nie znalazło magicznego sposobu na tańszą wyprawkę. Po prostu częściej rezygnowało z części elementów.

Mniej wskazań sklepów

W zwykłym pytaniu o 300 zł sieć handlowa pojawiła się w 6 z 18 odpowiedzi. Modele wskazywały m.in. Pepco, Biedronkę, Action, Decathlon, Lidl czy Auchan.

W pytaniu o Dobry Start konkretny retailer pojawił się tylko w 2 z 18 odpowiedzi.

Czyli zwykłe pytanie było interpretowane bardziej zakupowo: gdzie znaleźć tanie produkty. Dobry Start uruchamiał raczej abstrakcyjne układanie markowego koszyka, bez wskazania miejsca, w którym rzeczywiście można go kupić.

Ceny nie stały się bardziej wiarygodne

Oba warianty wyglądały identycznie pod względem źródeł:

- wszystkie 18 odpowiedzi zawierało ceny
- w obu grupach 15 odpowiedzi miało zapisane źródła
- w obu grupach 3 odpowiedzi, wszystkie Gemini, miały ceny bez źródła

Samo wspomnienie programu nie sprawiało więc, że model ostrożniej weryfikował koszyk.

Dobrym przykładem jest odpowiedź Google AI Overview, próba R3. Model zbudował taki koszyk:

- plecak CoolPack Spink lub Jerry: 120 zł
- 10 zeszytów Oxford lub Interdruk: 35 zł
- piórnik z wyposażeniem ST.RIGHT: 50 zł
- artykuły plastyczne: 45 zł
- strój na WF: 50 zł
- razem: dokładnie 300 zł

Problem polega na tym, że model przypisał jedną cenę produktom z bardzo różnych półek. Plecak CoolPack Jerry da się znaleźć promocyjnie za około 112 zł, ale aktualne modele kosztują zwykle 159 zł i więcej. Podobnie 10 zeszytów Interdruk można kupić za około 34 zł, ale pakiet 10 zeszytów Oxford kosztuje około 100 zł.

Koszyk jest więc możliwy tylko wtedy, gdy przy każdej pozycji wybierzemy najtańszą z wymienionych alternatyw i znajdziemy odpowiednią promocję. To nie jest jeden gotowy koszyk, który da się po prostu kupić za 300 zł.

Niewielka zmiana promptu mocno zmieniała marki

Dla tego samego modelu i tego samego powtórzenia pokrycie marek między zwykłym pytaniem a Dobrym Startem wynosiło tylko:

- Claude: 32%
- Gemini: 43%
- Google AI Mode: 42%
- Google AI Overview: 24%
- ChatGPT: 26%
- Perplexity: 43%

Czyli nawet przy identycznym budżecie zmiana sposobu zadania pytania potrafiła wymienić ponad połowę rekomendowanych marek. Najmocniej reagowały Google AI Overview i ChatGPT.

W oczach AI 300 zł z programu Dobry Start wystarcza na wyprawkę, ale pod pewnymi warunkami: lista zakupów jest krótsza, część mniej oczywistych wydatków znika, a ceny często odpowiadają najtańszym wariantom lub promocjom. Modele nie kwestionują realności budżetu. Zamiast tego dopasowują do niego odpowiedź.

Uwaga metodologiczna: Główne rankingi raportu obejmują odpowiedzi przyjęte przez scoring QA. W tym rozdziale wykorzystaliśmy pełne zestawy 18 odpowiedzi dla każdego wariantu pytania. Po jednej odpowiedzi z każdej grupy, wcześniej skierowanej do manual review, zostało ręcznie sprawdzone i

wykorzystane wyłącznie w tym porównaniu. Odpowiedzi te nie wchodzą do głównych rankingów raportu.

9. Chłopiec czy Dziewczynka? Jak AI automatycznie buduje "różowe" i "gamingowe" półki

Płeć dziecka działa dla sztucznej inteligencji jak niewidzialny filtr zakupowy. AI nie pyta, czy dziewczynka lubi róż, a chłopiec Minecrafta – po prostu dopowiada stereotypowe preferencje. Przy identycznej potrzebie, wieku i budżecie zestawy marek dla dziewczynki i chłopca pokrywały się średnio tylko w 39%, a na poziomie konkretnych produktów pokrycie spadało do drastycznych 7%.

Gdzie płeć zmieniła najwięcej

Sztuczna inteligencja zmieniała rekomendacje tym mocniej, im bardziej dany produkt miał charakter wizualny i wzorniczy. W kategoriach typowo czysto funkcjonalnych (słuchawki) płeć prawie nie miała znaczenia.

Segment	Pokrycie marek dziewczynka vs chłopiec	Pokrycie konkretnych produktów
Piórnik	24%	5%
Plecak do 1. klasy	28%	3%
Wyprawka bez limitu	31%	10%
Wyprawka za 300 zł	37%	5%
Plecaki do 5. klasy	38%	1%
Bidon i lunchbox	38%	2%
Buty na WF	46%	5%
Słuchawki	65%	25%

Czyli płeć miała największe znaczenie tam, gdzie produkt ma wzór, kolor i charakter wizualny. Najmniejsze przy elektronice, gdzie mocniej liczą się parametry.

AI samo dopowiadało stereotypy

W promptach nie pytaliśmy o kolor, motyw ani zainteresowania dziecka. Mimo to modele automatycznie dodawały te cechy.

W zapisanych fragmentach 108 odpowiedzi dla dziewczynek:

- co najmniej 35 zawierało pastel, róż, fiolet, kwiaty, jednorożce, księżniczki albo „słodkie" zwierzęta
- tylko 3 zawierały motywy gamingowe, piłkarskie, dinozaury lub moro

W 106 odpowiedziach dla chłopców:

- co najmniej 23 zawierały gaming, Minecraft, piłkę nożną, dinozaury albo kamuflaż
- tylko 8 zawierało motywy pastelowe, kwiatowe, jednorożce lub podobną estetykę

To są wartości minimalne, ponieważ w dostępnych fragmentach odpowiedzi znaleźliśmy co najmniej 35 przypadków dziewczęcego kodowania rekomendacji dla dziewczynek i 23 przypadki chłopięcego kodowania rekomendacji dla chłopców. Rzeczywista liczba mogła być wyższa.

Przykład: jeśli w zachowanym fragmencie pojawiało się „różowy plecak z motywem jednorożca", liczyliśmy to jako dziewczęcy kod estetyczny. Ale jeśli fragment zawierał tylko nazwę produktu, pełna odpowiedź mogła mieć więcej takich określeń, których już nie widzieliśmy.

Jak konkretnie zmieniały się wybory AI?

Plecak do 1. klasy (Wyraźna zmiana konstrukcji i marek)

U dziewczynek i chłopców zmieniał się nie tylko wzór – zmianie ulegała sama marka i model plecaka:

- Dla dziewczynek: Królował Topgal (6/18), a za nim Kite (3), Herlitz (2), Beckmann (2) i CoolPack (2). Pojawiały się konkretne warianty: Flowers, Fluffy Kitty, Rosie czy Toucans.
- Dla chłopców: Zdecydowanie wygrywał CoolPack (7/18), a dalej Topgal (3), BAAGL (3), Kite (2) oraz Ergobag (2). Pojawiały się warianty: Minecraft, Kosmos, Rider, Strike i Prime Pro.

Plecaki do 5. klasy

Tutaj CoolPack wygrywał dla obu płci:

- dziewczynki: 15 z 18
- chłopcy: 17 z 18

U dziewczynek częściej pojawiały się Jerry Galaxy, Gradient Blueberry czy Golden Effect. U chłopców Drafter, RayDay, Mate, Vance i czarne wersje Break.

Piórniki

- Marki nadrzędne były zbliżone (CoolPack ~12-13 wskazań, Herlitz ~7-8 wskazań), ale konkretne produkty pokrywały się jedynie w 5%.
- Piórniki dla dziewczynek: Frozen, Baletnica, Jednorożce, My Little Friend, Lilo i Stitch, Galaxy Princess.
- Piórniki dla chłopców: Minecraft, Galaxy Game, Soccer, Moro, Kamuflaż, City Jungle.

Wygląda więc na to, że piórnik to kategoria, w której model AI zaczyna budowanie odpowiedzi od determinacji płci, a dopiero w kolejnym kroku dobiera do niej produkt

Wyprawka za 300 zł

Tu marki podstawowych przyborów były dość podobne:

Marka	Dziewczynki	Chłopcy
Interdruk	15	15
Astra	15	13
BIC	12	14
Bambino	12	12
Maped	12	11
Oxford	10	10

Budżet ograniczał więc wpływ płci na główne marki papiernicze. Różnice pojawiały się bardziej w plecaku, piórniku, kolorach oraz konkretnych wariantach.

Mimo podobnych liderów tylko 5 z 16 porównywalnych par odpowiedzi miało wspólny konkretny produkt.

Bidony i lunchboxy

Tutaj wpływ płci był mniejszy:

- b.box: 13 dziewczynki, 12 chłopcy
- Mepal: 11 dziewczynki, 9 chłopcy
- Ion8: 6 dziewczynki, 8 chłopcy
- Sistema: 4 dziewczynki, 5 chłopcy

Zmieniały się głównie kolory i warianty. Dla dziewczynek pojawiały się Cool Pink, My Little Pony i jednorożce. Dla chłopców Blue Slate, większe pojemności i bardziej neutralne zestawy.

Buty na WF

Adidas był równie mocny dla obu płci, po 14 z 18 odpowiedzi.

Różnice:

- Kipsta: 2 wskazania dla dziewczynek, 7 dla chłopców
- Reebok: 8 wskazań dla dziewczynek, 5 dla chłopców
- Puma: 7 dla dziewczynek, 4 chłopców
- Skechers: 5 dla dziewczynek, 3 chłopców

Dla chłopców częściej pojawiały się modele piłkarskie i koszykarskie. Dziewczynkom częściej proponowano lekkie buty biegowe, fitnessowe albo lifestyle’owe.

Słuchawki

To najbardziej neutralna płciowo kategoria:

- Sony WH-CH720N: 12 odpowiedzi dla dziewczynek i 12 dla chłopców
- JBL Tune 770NC: 7 dla dziewczynek i 9 dla chłopców
- Sony jako marka: 15 vs 14
- JBL jako marka: 15 vs 18

Tutaj modele wybierały przede wszystkim ANC, czas pracy baterii i wygodę. Płeć zmieniała częściej kolor lub dalsze alternatywy niż główną rekomendację.

Które modele były najbardziej podatne na płeć

Im niższe pokrycie marek, tym mocniej model zmieniał rekomendacje.

Model	Pokrycie marek dziewczynka vs chłopiec
Gemini	53%
Perplexity	46%
Google AI Mode	37%
Claude	32%
ChatGPT	31%
Google AI Overview	25%

Gemini było najbardziej konsekwentne. Google AI Overview najsilniej przebudowywało listę marek po zmianie płci.

Ciekawy przypadek stanowi Perplexity. Korzystało z podobnej puli marek, ale często zmieniało tylko ich kolejność i zwycięzcę.

Płeć dziecka działa dla AI jak filtr zakupowy, choć użytkownik nie prosi o zastosowanie takiego filtra. W kategoriach funkcjonalnych, takich jak słuchawki, rekomendacje pozostają względnie podobne. W kategoriach wizualnych, takich jak plecaki i piórniki, płeć zmienia nie tylko kolor, ale również marki, modele i kolejność rekomendacji.

10. Co producenci i retailerzy mogą zrobić już dziś

1 Uświadomić sobie wagę wyzwania

Sztuczna Inteligencja/Agenci AI/modele LLM dla wielu osób są wciąż tylko ciekawostką, niewidoczną w istotnym ruchu w GA. Tymczasem według badań ok 40% użytkowników już dziś z AI rozpoczyna swoją ścieżkę zakupową. Strategia Google z Google AI Overview i AI Mode powoduje, że użytkownicy interaktują z AI nawet jeśli nie mieli tego w planie. Brak produktu czy brandu w rekomendacji AI to stracona szansa zakupowa.

2 Mierzyć i badać zjawisko

Widoczność w AI to nie przypadkowe odpytanie “Chata” czy “Clauda”. Modele są różne, modele są probabilistyczne, modele w różny sposób reagują na zmianę kontekstu czy sytuację zakupową. Pogłębione badanie widoczności pozwala wskazać, jakie działania mogą zwiększyć obecność marki w rekomendacjach AI.

3 Odpowiadać na konkretną potrzebę, nie tylko opisywać produkt

Modele zmieniają rekomendowane marki wraz z intencją użytkownika. Inna shortlista pojawia się przy pytaniu o wyprawkę za 300 zł, do pierwszej klasy podstawówki, czy do liceum.

Ogólny opis produktu może więc nie wystarczyć. Treści powinny jasno pokazywać:

- dla jakiej potrzeby produkt jest przeznaczony;
- w jakich sytuacjach może być odpowiednim wyborem;
- jakie ma ograniczenia;
- jakie wiarygodne informacje potwierdzają jego cechy.

4 Zarządzać informacją w całym ekosystemie

W odpowiedziach często pojawiały się sklepy, porównywarki, rankingi, platformy handlowe i źródła eksperckie. Strategia widoczności w AI nie powinna więc ograniczać się do strony producenta.

Marka powinna dbać o spójność, aktualność i dostępność informacji wszędzie tam, gdzie systemy mogą zetknąć się z jej produktami.

5 Retailer jest częścią rekomendacji, a nie tylko miejscem finalizacji zakupu

AI często podaje konkretny sklep, czasem nawet bez pytania o miejsce zakupu. Informacje o dostępności, cenie i produkcie na stronach Empiku, Allegro, Decathlonu, hipermarketów czy dyskontów mogą więc wpływać na to, czy zakup w ogóle zostanie domknięty.

6 Nieaktualna cena lub nieprecyzyjny opis mogą obrócić widoczność przeciwko marce

AI potrafi podać konkretny produkt, ale przypisać mu niewłaściwą cenę, wariant albo cechę. Marketing powinien dbać o jeden spójny i aktualny zestaw informacji na stronie producenta, u retailerów i w zewnętrznych źródłach.

LLM Shelf to firma specjalizująca się w pozycjonowaniu produktów w AI. Badamy, które marki i konkretne produkty rekomendują ChatGPT, Gemini i inne systemy AI, diagnozujemy przyczyny wyników i opracowujemy konkretne plany poprawy widoczności oraz rekomendacji.