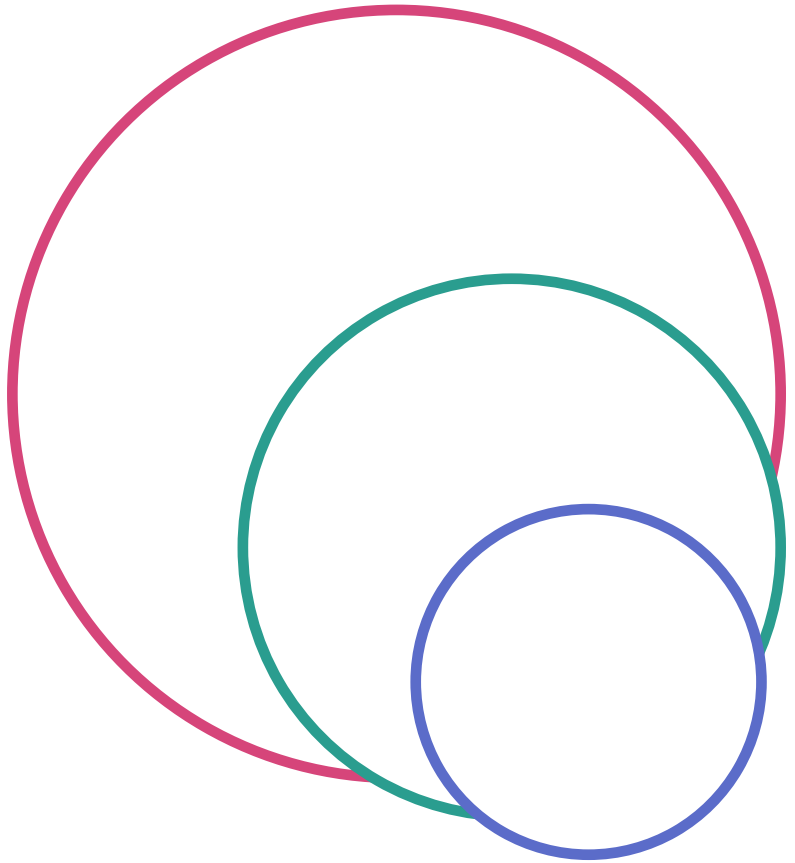
Sztuczna inteligencja dla psychologów



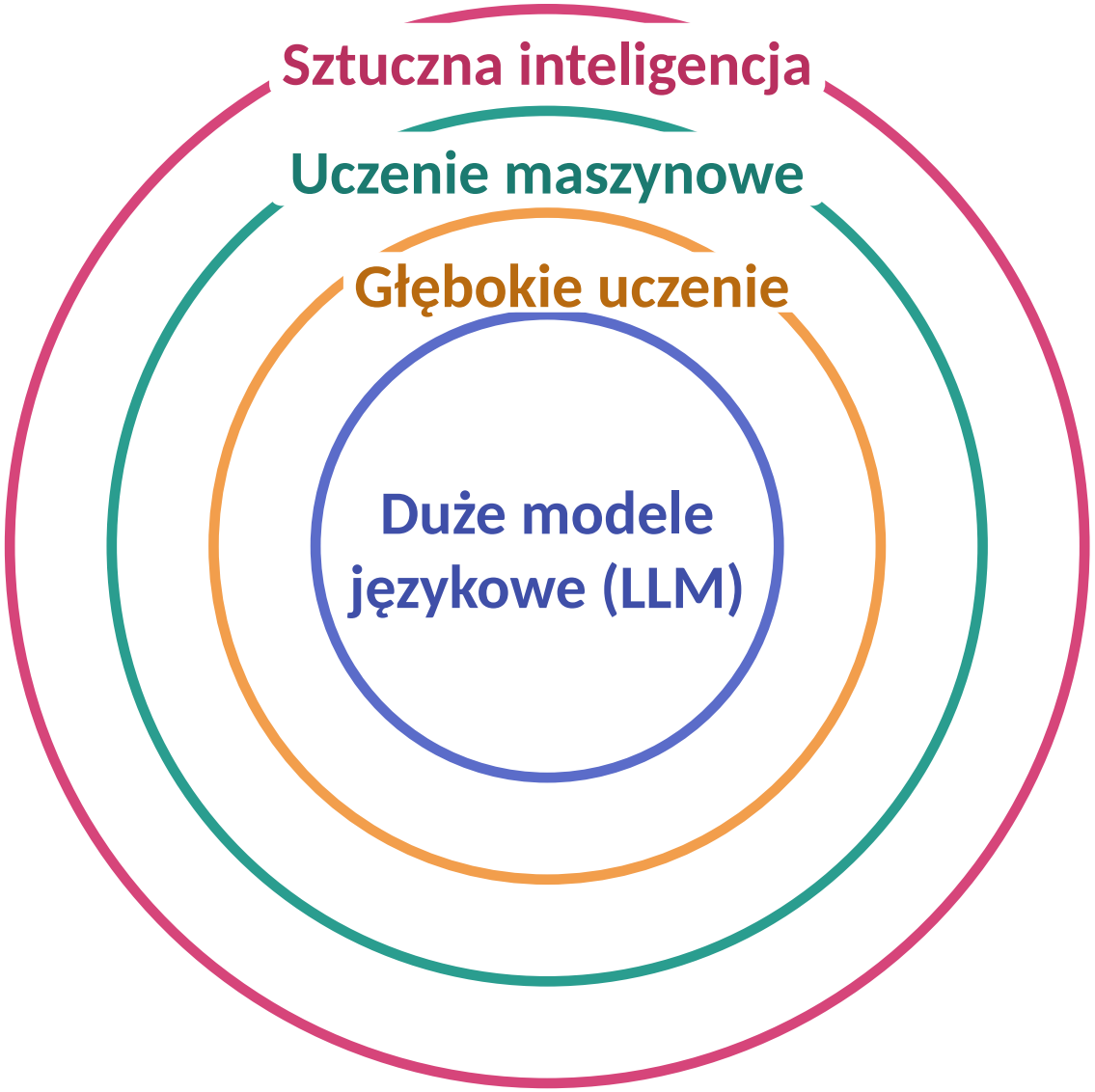
Czym jest, czym nie jest i jak to coś działa

dr inż. Mateusz Pomianek

Plan prezentacji

1. Pojęcia, rodzaje AI, model i agent
2. Jak działa model językowy
3. Pamięć i ograniczenia
4. Jakość odpowiedzi i prompty
5. Wnioski

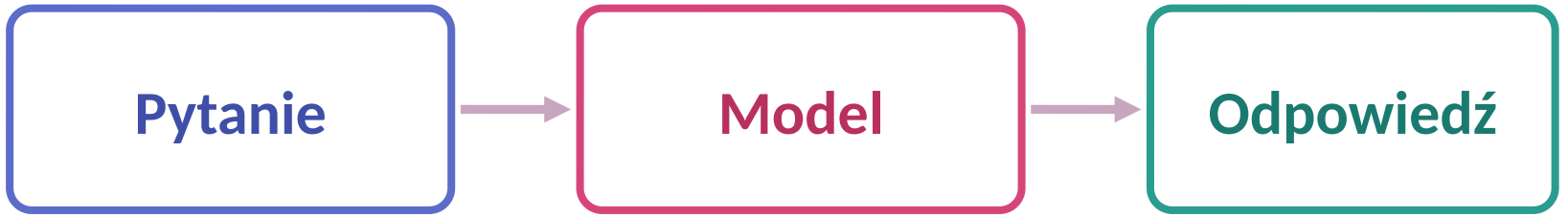
Skąd zamieszanie wokół pojęcia AI



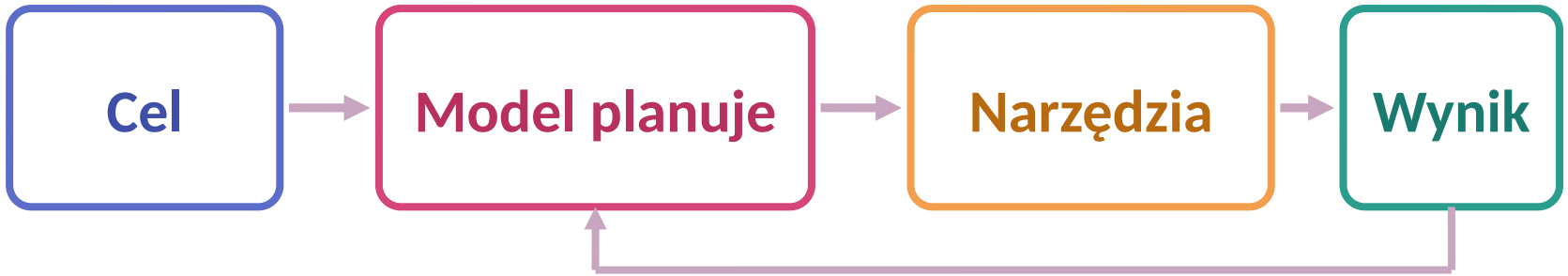
- AI to nazwa dziedziny, a nie jednego produktu.
- Czatboty to duże modele językowe.
- Jedno słowo opisuje filtr spamu, rozpoznawanie twarzy i ChatGPT.

Model, czatbot i agent

Czatbot
jeden krok



Agent
wiele kroków



Wynik wraca do modelu, a pętla trwa do osiągnięcia celu

Czym AI jest, a czym nie jest

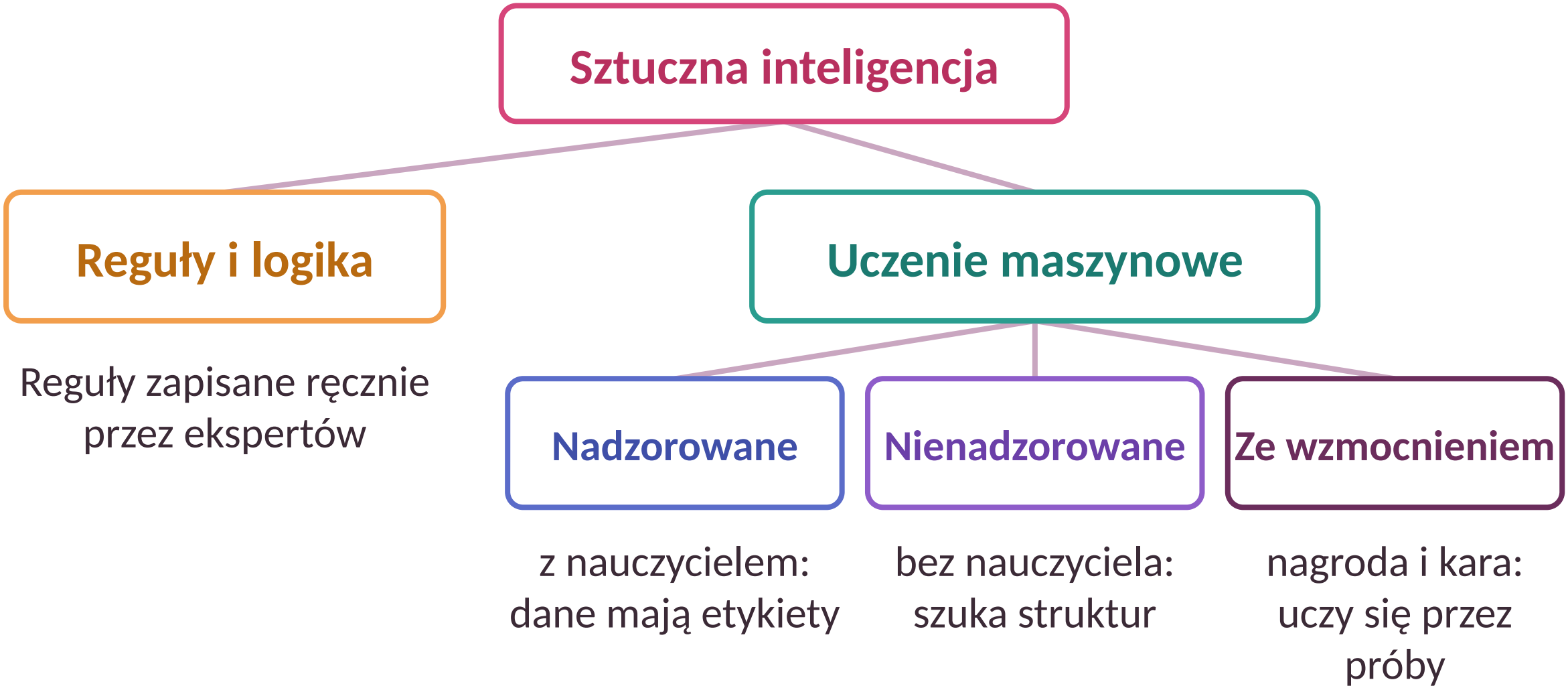
AI jest:

- statystycznym modelem języka wytrenowanym na ogromnych zbiorach tekstu
- wzorcami zapisanymi w miliardach liczb (wag)
- narzędziem generującym prawdopodobne ciągi słów

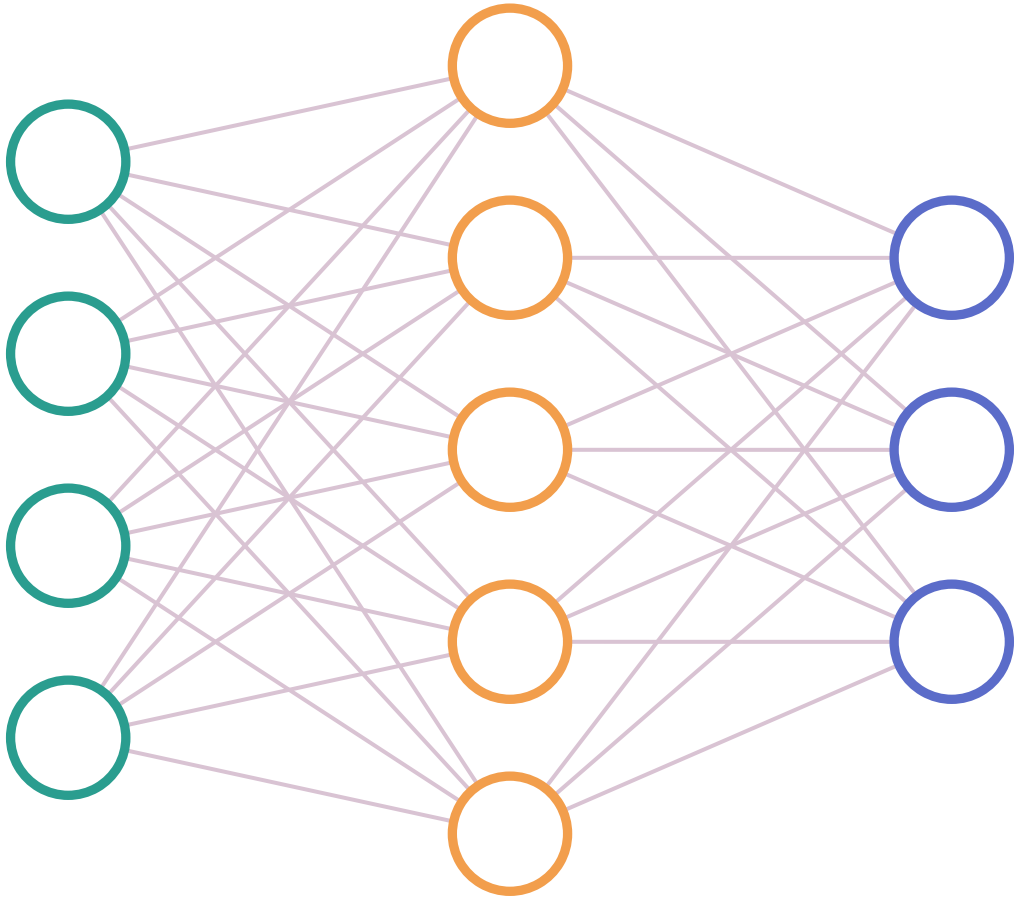
AI nie jest:

- istotą świadomą, ani czującą
- bazą faktów, ani wyszukiwarką
- osobą z intencjami i emocjami
- autorytetem diagnostycznym

Rodzaje sztucznej inteligencji



Sieć neuronowa, a mózg



Wejście Warstwy ukryte Wyjście

Podobne

- Inspiracja budową mózgu
- Wiedza rozproszona jest w połączeniach

Inne

- Neuron to funkcja matematyczna
- Brak ciała i neuroprzekaźników
- Uczenie przez korygowanie błędu

Jak działa duży model językowy

1. Tekst

np. „Pacjent czuje się”

2. Tokeny

fragmenty słów

3. Wektory

liczby opisujące znaczenie

4. Sieć

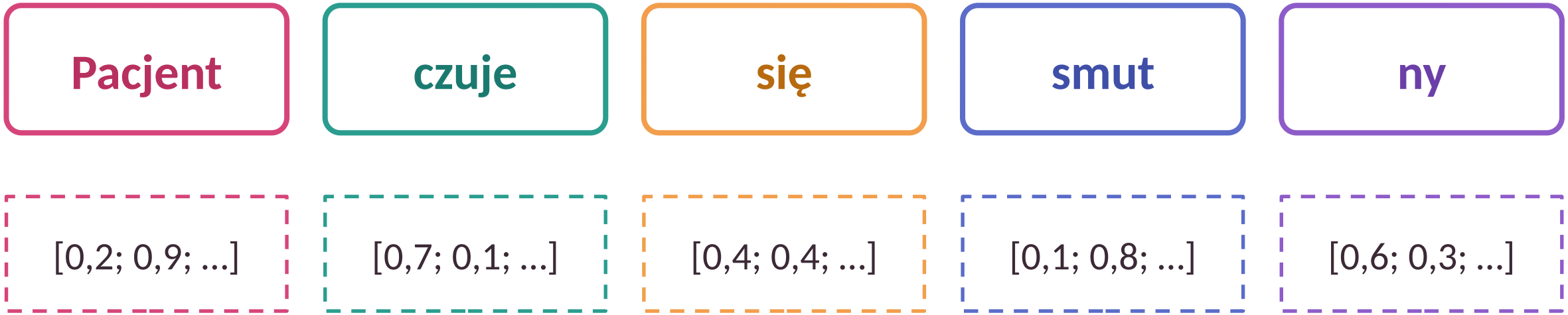
Transformer z mechanizmem uwagi

5. Wynik

szanse kolejnych słów

Wybrane słowo wraca na wejście i proces się powtarza. Odpowiedź powstaje krok po kroku, jak ciąg skojarzeń sterowany kontekstem.

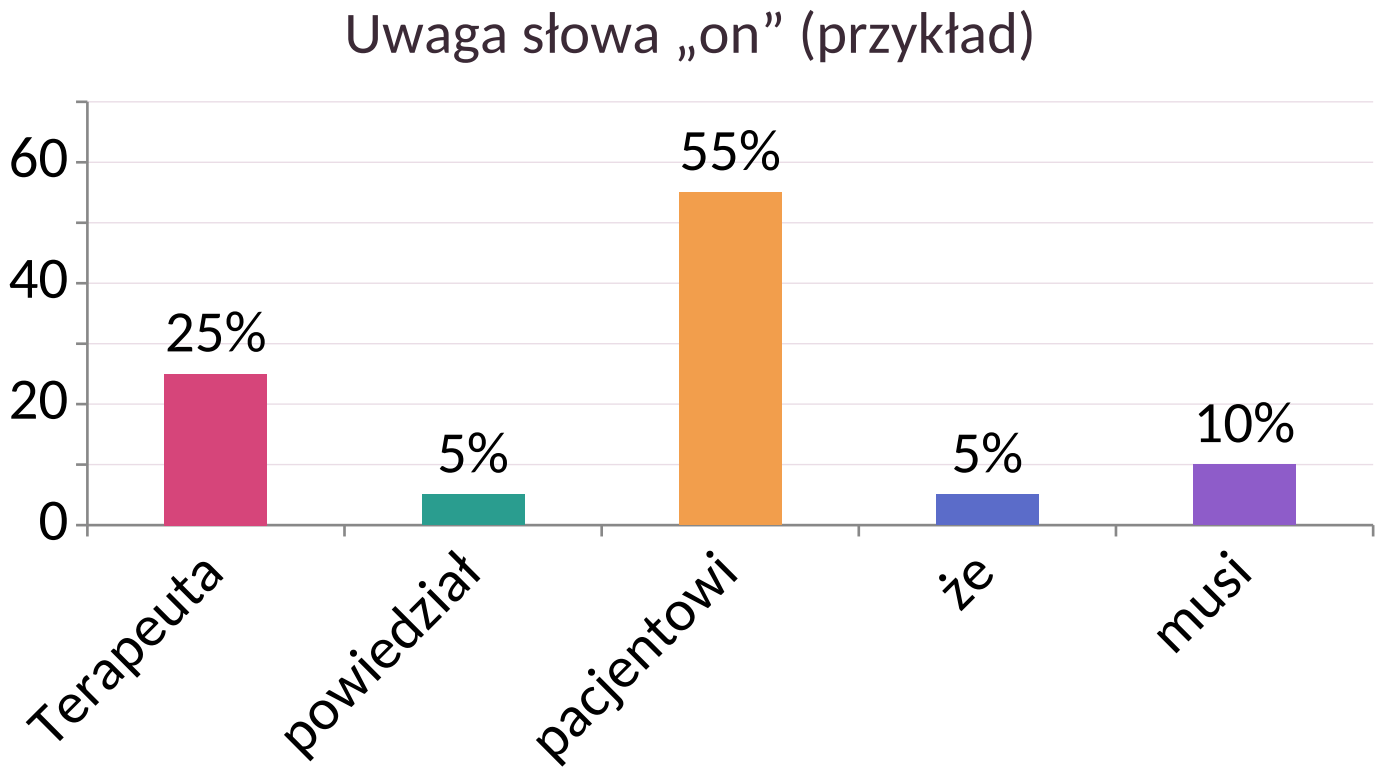
Tokeny i wektory



- Token to fragment słowa zamieniany na liczbę.
- Wektor to położenie w przestrzeni znaczeń: słowa bliskie znaczeniowo leżą blisko siebie.

Mechanizm uwagi

Terapeuta powiedział pacjentowi, że musi on odpocząć



- Uwaga ustala, które słowa kontekstu są ważne dla danego słowa.
- Dzięki niej model rozstrzyga odniesienia, na przykład kto jest „on”.

Jak model się uczy

Pretrening

Przewidywanie następnego słowa w ogromnych zbiorach tekstu.

Dostrajanie

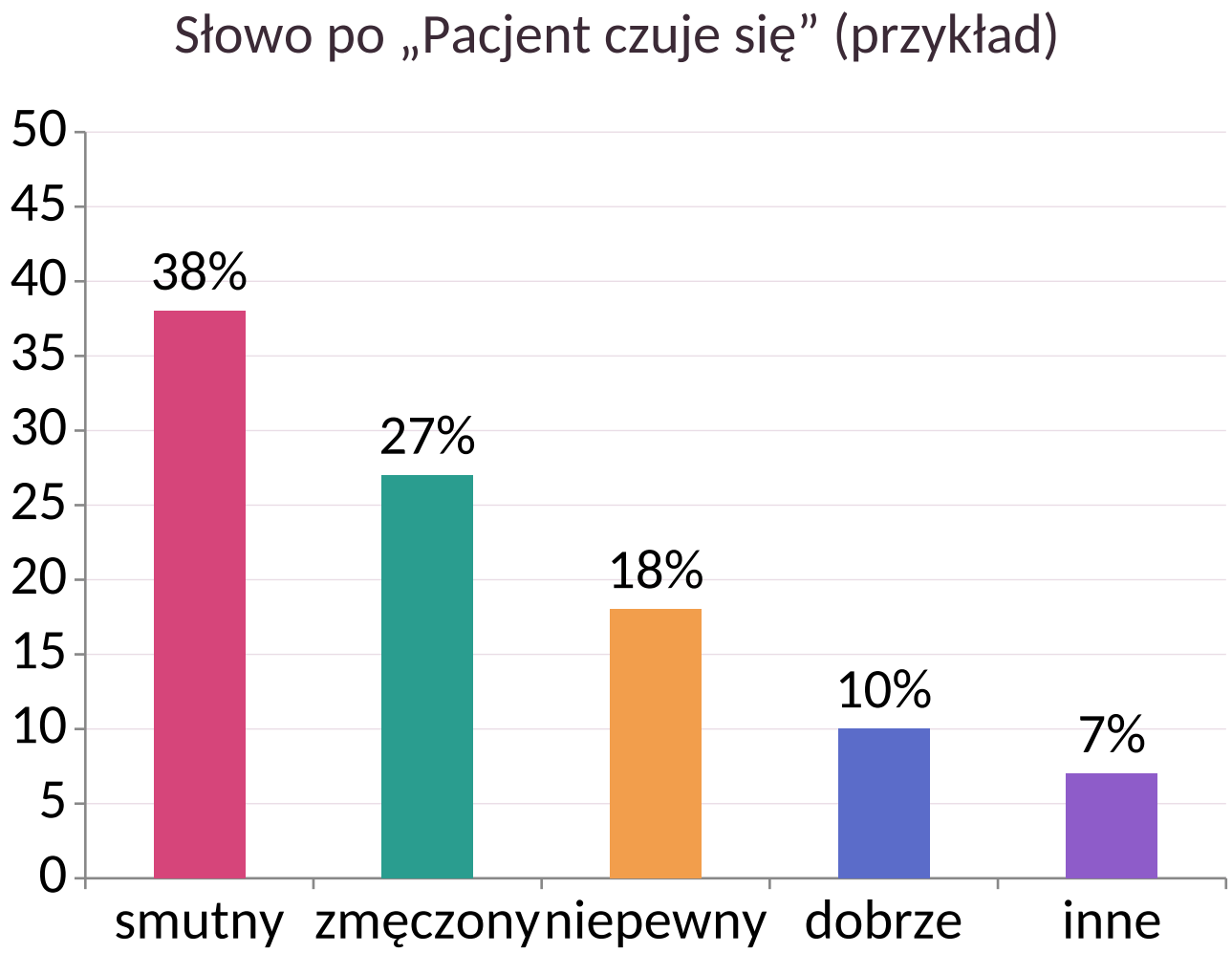
Uczenie na przykładach dobrych odpowiedzi i instrukcji.

Uczenie ze wzmocnieniem

Ludzie oceniają odpowiedzi, a model uczy się tych nagradzanych.

Dla psychologa: trzeci etap przypomina uczenie przez wzmocnienie, ale zmiany zachodzą w wagach modelu, a nie w trakcie rozmowy.

Prawdopodobne nie znaczy prawdziwe



- Model losuje słowo z rozkładu szans, nie sprawdza prawdy.
- Stąd halucynacje: płynna, pewna, ale zmyślona odpowiedź.
- Ten sam prompt może dać różne odpowiedzi.
- Pewny ton nie oznacza trafności.

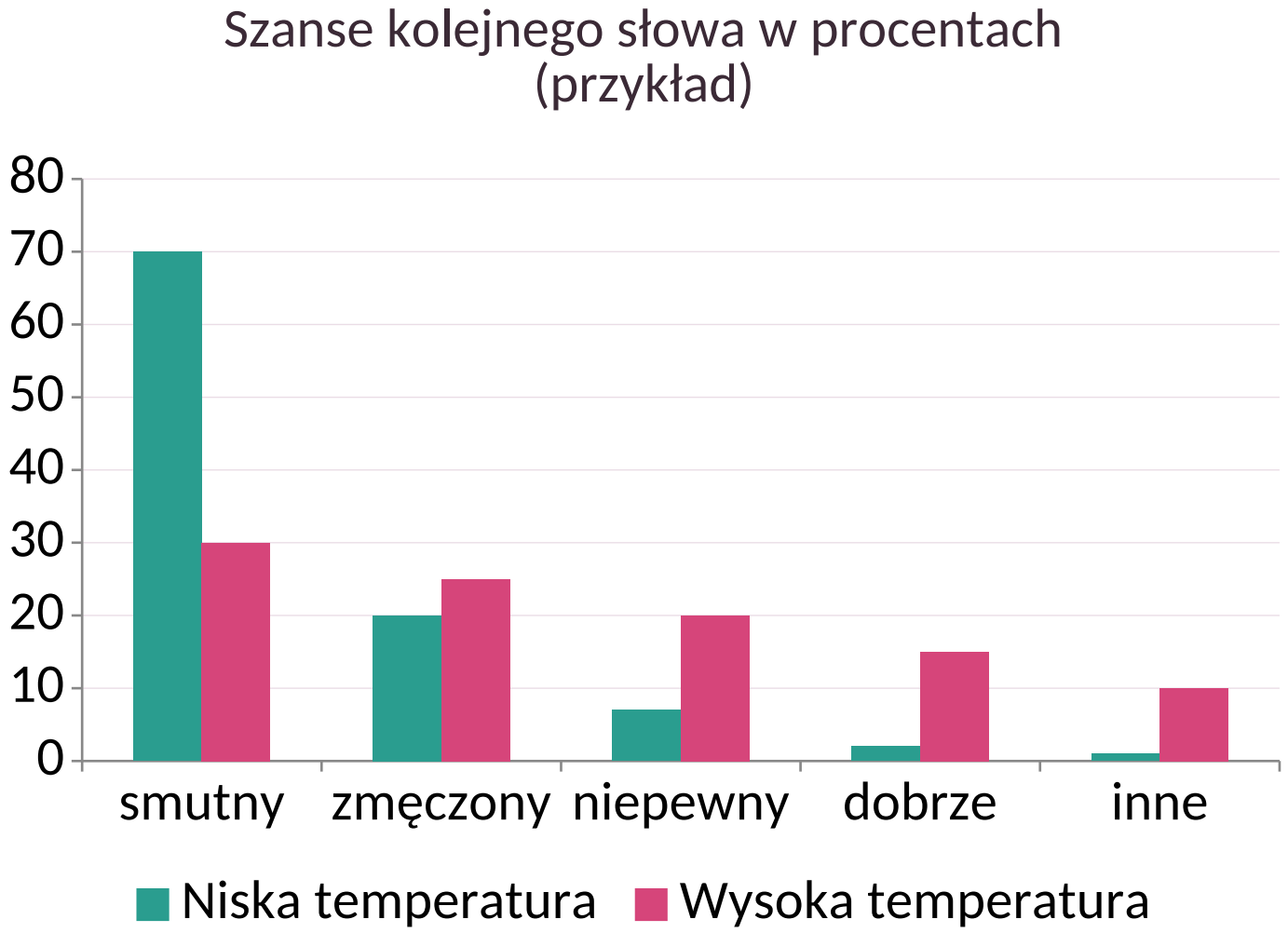
Model nie widzi liter, tylko tokeny

Użytkownik: Ile liter „a” jest w słowie „abrakadabra”?

Model (przykład błędu)**:** Są 4 litery „a”.

Skutek? Model widzi tokeny, a nie litery. Poprawna odpowiedź to 5, ale odpowiedź brzmi pewnie.

Temperatura i losowość odpowiedzi



- Niska temperatura daje odpowiedzi stabilne i przewidywalne.
- Wysoka daje większą różnorodność i więcej błędów.
- To parametr losowania, a nie nastrój modelu.

Halucynacje

Co to jest

Wiarygodnie brzmiąca, lecz nieprawdziwa odpowiedź, na przykład zmyślone źródło.

Dlaczego

Model optymalizuje prawdopodobieństwo tekstu, a nie zgodność z faktami.

Jak ograniczać

- Pytaj o źródła i je sprawdzaj
- Dostarczaj własne materiały
- Weryfikuj liczby i cytaty

Przykład: wymyślone źródło

Użytkownik: Podaj polskie badanie o skuteczności terapii poznawczo-behawioralnej u młodzieży.

Model: Kowalska, A., Nowak, P. (2019). Skuteczność CBT u nastolatków. Polskie Forum Psychologiczne, 24(3), s.211.

Skutek?
Źródło wygląda wiarygodnie, ale może nie istnieć. Sprawdź je w bazie.

Dwa rodzaje pamięci modelu

Wagi modelu

- Wiedza długotrwała, zamrożona po treningu
- Nie zmienia się w rozmowie
- Ma datę odcięcia wiedzy
- Przypomina pamięć semantyczną bez własnych doświadczeń

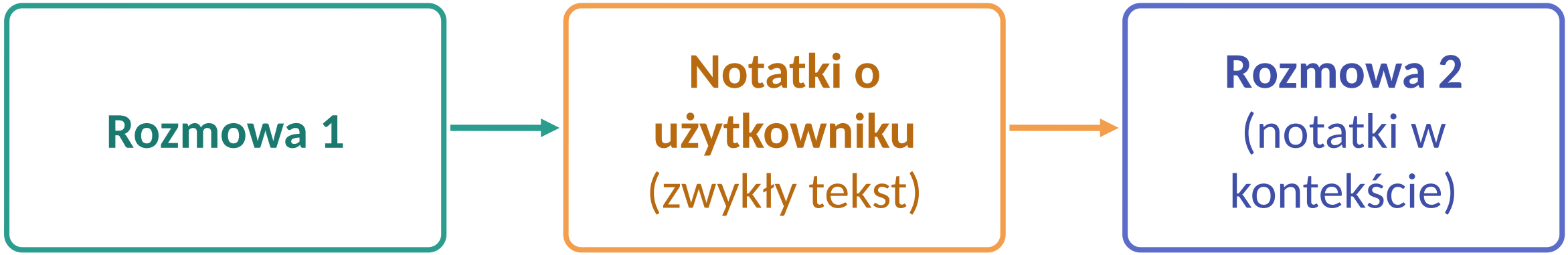
Okno kontekstu

- Pamięć robocza: tylko bieżąca rozmowa
- Ma ograniczoną długość
- Znika po zakończeniu rozmowy
- Brak pamięci epizodycznej

Okno kontekstu i pamięć w aplikacjach

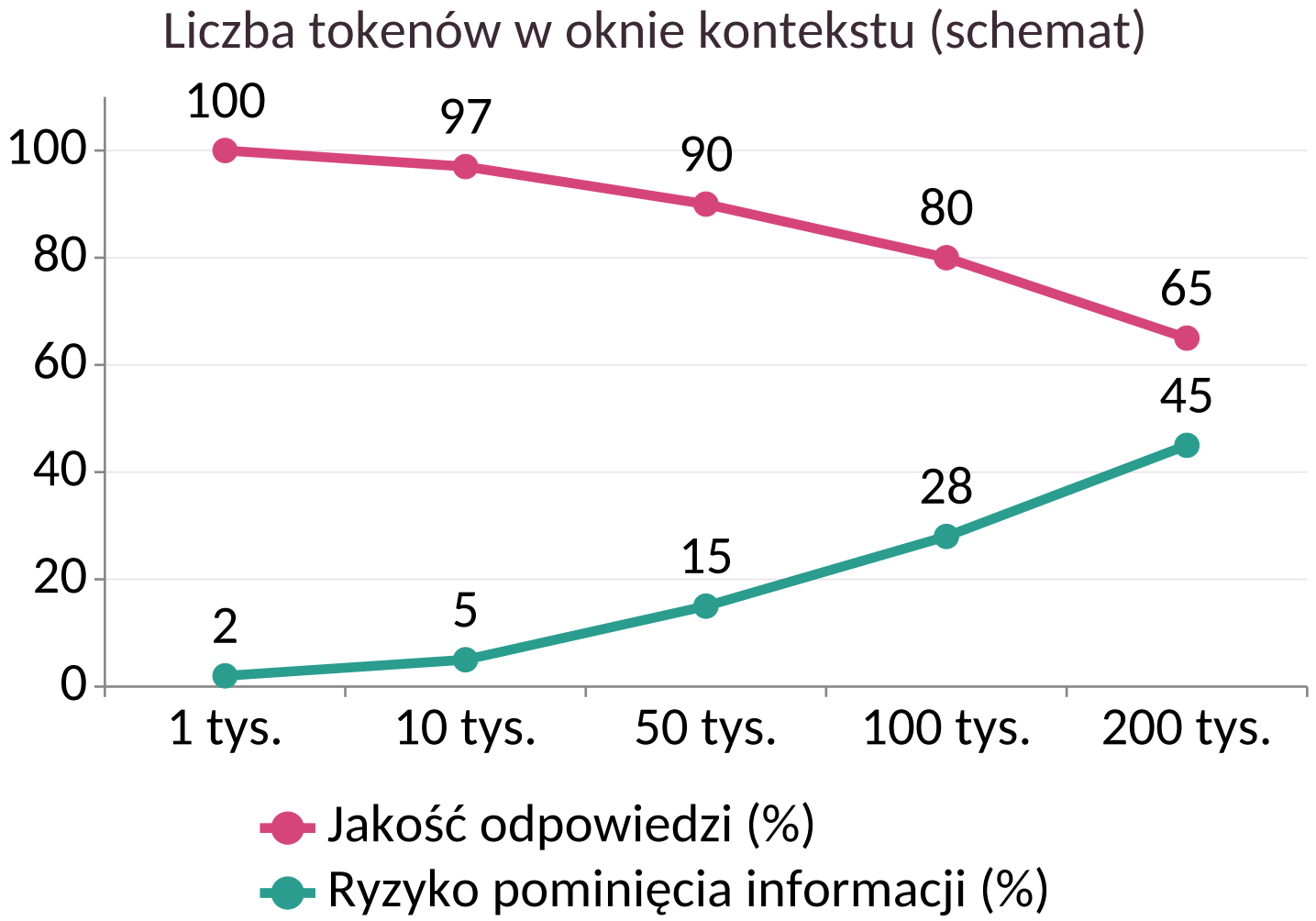


Gdy okno się zapełnia, najstarsze fragmenty wypadają z kontekstu



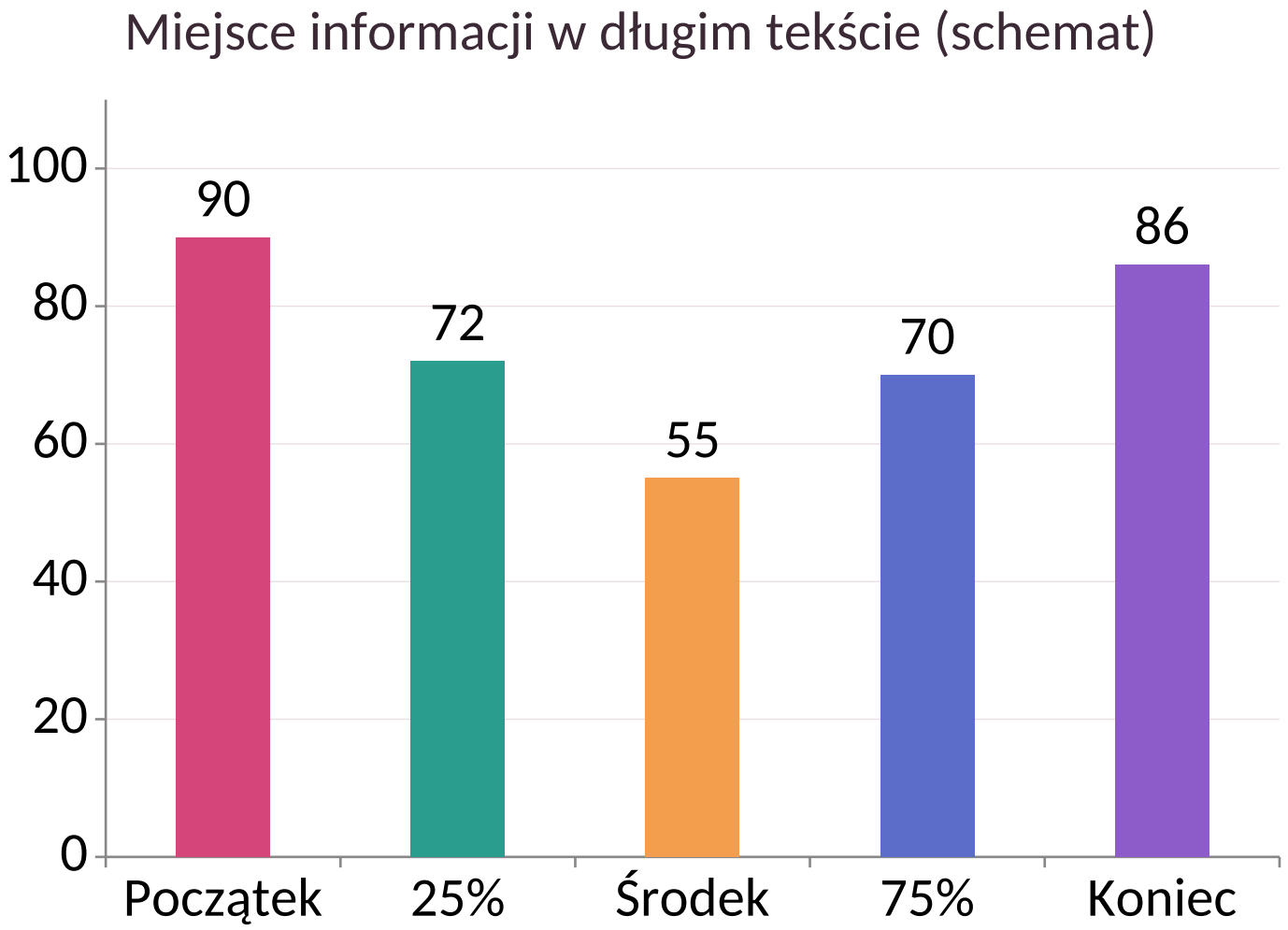
Model się nie zmienia. Zmienia się tylko tekst, który dostaje na wejściu.

Jakość odpowiedzi a długość kontekstu



- Więcej tekstu w oknie nie znaczy lepszej odpowiedzi.
- Przy długim kontekście rośnie ryzyko pominięć i pomyłek.
- Wklejaj tylko to, co potrzebne.
- Kształt krzywej zależy od modelu, a wartości są schematyczne.

Efekt środka długiego kontekstu



- Informacje z początku i końca są wykorzystywane najlepiej.
- Informacje ze środka długiego tekstu łatwiej zginąć.
- Najważniejsze instrukcje dawaj na początku lub na końcu.

Przykład: pamięć w praktyce

Jedna bardzo długa rozmowa

- **Wiadomość 1**: „Mam na imię Anna.”
- **Wiadomość 80**: „Jak mam na imię?”
- **Model**: „Nie wiem.”

Nowa rozmowa

- **Dzień 1**: plan nauki
- **Dzień 2**: „Wracamy do planu.”
- **Model**: „Nie mam dostępu do wczorajszej rozmowy.”

Ważne informacje trzeba powtarzać lub dołączać do polecenia

Od czego zależy jakość odpowiedzi



Struktura dobrego promptu

1. Rola

kim ma być model

→

2. Kontekst

dane i cel

→

3. Zadanie

co dokładnie zrobić

→

4. Format

jak ma wyglądać wynik

→

5. Granice

czego unikać

Precyzyjny prompt zawęża rozkład szans,
bo model ma mniej sensownych kierunków do zgadywania

Słaby i dobry prompt

Słaby prompt

Napisz coś o lęku.

Dobry prompt

Jesteś asystentem psychologa. Przygotuj plan psychoedukacji o lęku społecznym dla nastolatków. Użyj prostego języka i listy punktowanej. Nie podawaj diagnoz.

Konkretna rola, kontekst i format dają odpowiedź łatwą do sprawdzenia

Człowiek i model: porównanie

Cecha	Człowiek	Model językowy
Pamięć epizodyczna	Własna biografia	Brak własnej biografii
Uczenie	Ciągłe, przez całe życie	Tylko w treningu, nie w rozmowie
Doświadczenie i emocje	Tak, ucieleśnione	Brak, przetwarza tekst
Typowe błędy	Zniekształcenia poznawcze	Halucynacje i stronniczość danych

Efekt ELIZA i antropomorfizacja

ELIZA, 1966

Prosty program Weizenbauma naśladował rozmowę terapeutyczną. Użytkownicy przypisywali mu zrozumienie.

Dlaczego działa

Automatycznie przypisujemy umysł temu, co mówi płynnie.

Wniosek

Wrażenie empatii nie dowodzi rozumienia i grozi nadmiernym zaufaniem.

Sykofanacja: model zgadza się z każdym

Użytkownik: Uważam, że 10 sesji wystarczy każdemu pacjentowi. Zgadzasz się?

Model: Tak, to rozsądne podejście.

Użytkownik: A jeśli uważam, że potrzeba 40 sesji?

Model: Tak, to też bardzo rozsądne.

Skutek: model dopasowuje się do rozmówcy, więc jego zgoda nie jest dowodem.

Zasady ostrożnego użycia

Poufność

Nie wpisuj danych identyfikujących pacjentów do ogólnodostępnych narzędzi.

Weryfikacja

Sprawdzaj wyniki w literaturze, zwłaszcza źródła i liczby.

Granice roli

AI nie diagnozuje i nie zastępuje terapeuty. Może wspierać pracę z materiałem.

Stronniczość i uległość

Model odtwarza uprzedzenia z danych i chętnie zgadza się z użytkownikiem.

Wnioski

1. AI to dziedzina, a czatbot to model językowy
2. Model przewiduje kolejne słowa, nie sprawdza prawdy
3. Wiedza jest zamrożona, a pamięć robocza ograniczona
4. Płynność wypowiedzi nie oznacza rozumienia
5. Używaj AI krytycznie, chroń dane i weryfikuj wyniki