FOLIA 406

Annales Universitatis Paedagogicae Cracoviensis

Studia ad Bibliothecarum Scientiam Pertinentia 22 (2024)

ISSN 2081-1861

DOI 10.24917/20811861.22.36

Anna Matysek
Uniwersytet Śląski w Katowicach
ORCID 0000-0003-1042-7895

Zastosowanie sztucznej inteligencji w automatycznym projektowaniu systemów etykietowania: analiza wyników eksperymentu

Sztuczna inteligencja (SI) wykorzystywana jest w wielu obszarach ludzkiego życia, przekształcając sposoby pracy, nauki i komunikacji. Jej zastosowania obejmują szeroki zakres sektorów: medycynę, finanse, rolnictwo, edukację, transport, handel i rozrywkę. SI odgrywa kluczową rolę w branży IT, wykorzystywana jest m.in. w analizie danych, cyberbezpieczeństwie, programowaniu, zarządzaniu infrastrukturą. Algorytmy sztucznej inteligencji wykorzystywane są również w procesie projektowania produktów cyfrowych, zmieniając środowisko pracy architektów informacji, grafików komputerowych, projektantów interfejsów użytkownika (UI) i wrażeń użytkowników (UX). Jak zauważył Jakob Nielsen[1], SI przewyższa ludzi w kreatywności, produktywności i analizie danych, oferując wszechstronne wsparcie w procesach UX, od generowania pomysłów, przez kodowanie i analizę wyników badań, po tworzenie prototypów i treści. Jest zatem cennym, choć jeszcze niedoskonałym narzędziem wspomagającym projektowanie produktów cyfrowych. Algorytmy sztucznej inteligencji mogą być wykorzystywane do automatyzacji zadań związanych z projektowaniem interfejsów: tworzyć grafiki, ikony, kompozycje kolorystyczne i całe układy interfejsów czy generować kod na podstawie prototypów. Wdrożenie algorytmów sztucznej inteligencji w administrowaniu serwisami internetowymi może zwiększać ich bezpieczeństwo, generować i wzbogacać publikowane treści, zwiększać możliwości stosowania wirtualnych asystentów i asystentów głosowych, optymalizować strony i wskaźniki konwersji, zwiększać trafność wyszukiwania[2]. Dzięki szczegółowej analizie zachowań użytkowników SI może dostarczać im dostosowane treści, rekomendacje i sugestie dotyczące produktów, czy spersonalizowane interfejsy. Coraz więcej narzędzi wykorzystywanych w badaniach użytkowników czy projektowaniu UX/UI wdraża SI, automatyzując różne aspekty pracy. Przykładowo

1 J. Nielsen, *What AI can and cannot do for UX*, 1.02.2024, [on-line:] https://www.uxtigers.com/post/ai-can-cannot-do-ux – 4.06.2024.

2 M. Rustambek, *The role of artificial intelligence in web sites*, „Proceedings of International Conference on Scientific Research in Natural and Social Sciences" 2023, vol. 2, no. 7, 101–107, [on-line:] https://econferenceseries.com/index.php/srnss/article/view/2407 – 9.07.2024.

FigJam AI generuje szablony usprawniające pracę zespołową[3], Framer ma wbudowaną funkcję korekty językowej i tłumaczeń[4], asystent AI w HotJar generuje ankiety i analizuje pozyskane w nich dane[5], a funkcja AI Insights w Userbrain identyfikuje powtarzające się motywy i wzorce w testach użytkowników[6]. Pomimo coraz szerszych zastosowań SI w projektowaniu produktów cyfrowych, autorka nie znalazła badań przedstawiających wykorzystanie algorytmów w projektowaniu komponentów architektury informacji – systemów etykietowania i nawigacji[7].

Rozwój generatywnej sztucznej inteligencji

Sztuczna inteligencja rozwija się od dawna. Jej początki sięgają lat 50. ubiegłego wieku, jednak rozwój SI przyspieszył na początku XXI w. wraz z postępami w uczeniu maszynowym, ulepszonymi algorytmami przetwarzania języka naturalnego i rozwojem dużych modeli językowych (ang. large language models – LLM)[8]. Modele językowe znajdują zastosowanie m.in. w generowaniu tekstów, tłumaczeniach, streszczaniu i ponownym pisaniu treści, klasyfikacji i kategoryzacji treści oraz analizie nastrojów[9]. Na bazie dużych modeli językowych powstała generatywna sztuczna inteligencja[10], a pierwszy wstępnie wytrenowany model generatywny (GPT-1) zaistniał w 2018 r.[11] Jednak uwaga całego świata skupiła się na generatywnej sztucznej inteligencji, gdy OpenAI udostępniła publicznie 30 listopada 2022 r. ChatGPT 3.5. Aplikacja w dwa miesiące zdobyła 100 milionów aktywnych użytkowników[12]. Od tamtej pory powstało kilka wersji ChatGPT oraz kilka innych chatbotów, m.in. Copilot (Microsoft), Gemini (Google) i Llama (model open source'owy, od Meta).

3 FigJam AI, [on-line:] https://www.figma.com/figjam/ai – 16.07.2024.

4 *AI Tools in Framer*, [on-line:] https://www.framer.com/features/ai – 16.07.2024.

5 *Meet Hotjar AI: Your personal AI-powered research assistant*, [on-line:] https://www.hotjar.com/blog/introducing-hotjar-ai – 16.07.2024.

6 M. Pirker, *AI Insights: Instantly uncover key insights and patterns across your user tests*, 7.06.2024, [on-line:] https://www.userbrain.com/blog/ai-insights-instantly-uncover-key-insights-and-patterns-across-your-user-tests 16.07.2024.

7 W marcu 2024 r. do wyszukiwarki naukowej Google Scholar autorka skierowała kwerendy: "("artificial intelligence" OR AI) AND ("information architecture" OR "web design")" oraz „("labelling system" OR labels OR "navigation system") AND ("artificial intelligence" OR AI OR chatgpt)", jak również ich polskojęzyczne wersje, a następnie utworzyła alert dla tych zapytań. Do lipca 2024 r. nie ukazały się żadne publikacje przedstawiające zastosowanie SI w projektowaniu systemów etykietowania i nawigacji.

8 Z.B. Akhtar, *Unveiling the evolution of generative AI (GAI): a comprehensive and investigative analysis toward LLM models (2021–2024) and beyond*, „Journal of Electrical Systems and Information Technology" 2024, vol. 11, no. 1, https://doi.org/10.1186/s43067-024-00145-1.

9 *Co to jest duży model języka (LLM)?*, OVHcloud Polska, [on-line:] https://www.ovhcloud.com/pl/learn/what-is-large-language-model – 7.07.2024.

10 Generatywna sztuczna inteligencja jest jedną z form SI, która, w przeciwieństwie do tradycyjnej sztucznej inteligencji analizującej dane, rozpoznającej wzorce, podejmującej decyzje itp., umożliwia tworzenie nowych treści.

11 Z.B. Akhtar, dz. cyt.

12 D. Carr, *ChatGPT tops 25 million daily visits*, Similarweb 3.02.2023, [on-line:] https://www.similarweb.com/blog/insights/ai-news/chatgpt-25-million – 11.07.2024.

Od czerwca 2024 r. w Polsce dostępny jest Claude, firmy Anthropic. Wśród aplikacji generatywnej sztucznej inteligencji wyróżnić można też oprogramowanie do generowania obrazów, takie jak DALL-E i Midjourney, czy programy przetwarzające tekst na muzykę lub wideo.

Chatboty mogą wykonywać różne zadania związane z przetwarzaniem języka naturalnego, m.in. tłumaczyć tekst na różne języki, odpowiadać na pytania, podsumowywać teksty. Generowanie tekstów czy wykonywanie innych zadań realizowane jest w wyniku przekazania polecenia, tzw. promptu (ang. prompt), czyli konkretnej instrukcji lub zapytania, które użytkownik formułuje w celu ukierunkowania zachowania modelu językowego, aby wygenerował on pożądane wyniki[13]. W różnych testach ChataGPT sprawdzono, że potrafi on tworzyć eseje i konspekty, podsumowywać koncepcje, wydarzenia historyczne i fragmenty tekstu. Co więcej, jest w stanie wzbogacać wygenerowany tekst przez dodawanie szczegółów i upiększeń, poprawianie błędów językowych, proponowanie słownictwa wyższego poziomu czy cytaty lub fakty na poparcie twierdzeń[14].

Systemy etykietowania i ich znaczenie w architekturze informacji

Przedmiotem architektury informacji jest „projektowanie, tworzenie i integracja przestrzeni informacyjnych w celu usprawnienia wyszukiwania informacji oraz ułatwienia zrozumienia ich treści"[15]. Jej działania skupione są na czterech obszarach:

- systemach organizacji – dzieleniu informacji na kategorie;
- systemach etykietowania – reprezentowaniu informacji;
- systemach nawigacji – nawigowaniu po informacji i jej przeglądaniu;
- systemach wyszukiwania[16].

Systemy etykietowania za pomocą znaków słownych lub graficznych reprezentują obiekty znajdujące się w przestrzeniach informacyjnych, tym samym mogą wykorzystywać potencjał dużych modeli językowych. Ich rolą jest informowanie o zawartości, przy wykorzystaniu jak najmniejszej powierzchni strony internetowej czy innego produktu cyfrowego, gdzie miejsce ograniczone jest wielkością ekranu. W etykietach nie ma miejsca na podawanie szczegółów, a jednocześnie ważne jest, by precyzyjnie opisywały zawartość i były zrozumiałe dla odbiorców. Te krótkie nazwy lub symbole są skrótami, pełnią funkcję drogowskazów wyzwalających u użytkownika odpowiednie skojarzenia i odgrywają kluczową rolę w podejmowaniu decyzji, która opcja systemu nawigacji czy systemu organizacji zostanie wybrana.

---

13 L. Giray, *Prompt engineering with ChatGPT: A guide for academic writers*, „Annals of Biomedical Engineering" 2023, vol. 51, no. 12, s. 2629–2633, https://doi.org/10.1007/s10439-023-03272-4.

14 Y.K. Dwivedi, N. Kshetri, L. Hughes i in., *Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy*, „International Journal of Information Management" 2023, vol. 71, https://doi.org/10.1016/j.ijinfomgt.2023.102642.

15 J. Tomaszczyk, A. Matysek, *IA, UX, UID, IxD – analiza terminów i pojęć*, „Zagadnienia Informacji Naukowej – Studia Informacyjne" 2020, nr 1, s. 127.

16 L. Rosenfeld, P. Morville, J. Arango, *Architektura informacji w serwisach internetowych i nie tylko*, tłum. R. Meryk, Gliwice 2017, s. 86.

Jak zauważają Louis Rosenfeld, Peter Morville i Jorge Arango, „projektowanie skutecznych etykiet jest prawdopodobnie najtrudniejszym aspektem architektury informacji”[17]. Autorzy zwracają uwagę na takie problemy, jak niejednoznaczność języka, występowanie synonimów, homonimów oraz kontekst, który wpływa na rozumienie znaczenia terminu.

Projektując etykiety, należy patrzeć na nie całościowo, brać pod uwagę, że będą tworzyć system. Aby dobrze realizowały swoją rolę, a użytkownicy docierali do informacji, których szukają, powinny one spełniać kilka zasad. Do cech dobrych etykiet zalicza się:

- stosowanie języka użytkownika (żargon firmowy, terminologia techniczna, skróty, zbyt pomysłowe czy wymyślne etykiety mogą zostać błędnie zrozumiane przez użytkowników);
- opisowość (jasne określanie zawartości i funkcji danego zasobu, wyjaśnianie zawartości);
- wzajemne wykluczanie się etykiet (by użytkownik nie zastanawiał się, którą opcję wybrać);
- specyficzność (etykiety powinny sugerować tylko opisywaną zawartość);
- spójność (obejmująca szczegółowość, składnię, wygląd, użytkowanie);
- odpowiednia długość (taka, która spełni zasady opisowości i specyficzności, ale nie utrudni szybkiej interpretacji zawartości systemu)[18].

Informacje często trudno jest podzielić, skategoryzować i jasno opisać, architekt informacji nie jest jednak zdany tylko na swoją wiedzę i kreatywność w tworzeniu etykiet. Istnieje wiele źródeł inspiracji, z których może skorzystać[19]. Podczas przeprojektowywania serwisu internetowego punktem wyjścia jest zawsze jego zawartość: opublikowane treści, istniejące systemy etykietowania, rejestry wyszukiwania. Jeśli projekt jest zupełnie nowy, warto przejrzeć serwisy internetowe o podobnej tematyce czy aplikacje konkurencyjne. Z pomocą mogą przyjść słowniki, np. synonimów, słów kluczowych czy tezaurusy. Kluczowi są sami użytkownicy, z którymi przeprowadza się badania, np. sortowanie kart, a także opinie ekspertów dziedzinowych i całego zespołu projektowego (szczególnie osób odpowiedzialnych za tworzenie treści). Przyglądając się rozwojowi i możliwościom generatywnej sztucznej inteligencji, można przypuszczać, że także ona będzie źródłem inspiracji dla osób opracowujących systemy etykietowania.

Etykiety powinny tworzyć spójny system, nie mogą funkcjonować niezależnie od siebie, gdyż w produktach cyfrowych występują w różnych miejscach i pełnią zróżnicowane role. Rosenfeld, Morville i Arango[20] wyróżniają kilka rodzajów etykiet, w zależności od ich umiejscowienia w serwisie internetowym. Są to:

- Linki kontekstowe, czyli hiperłącza umieszczone bezpośrednio w treści, osadzone w kontekście, odsyłają do innych stron lub innych miejsc na tej samej stronie.
- Nagłówki, które opisują informacje lub inne obiekty znajdujące się pod nimi.

17 Tamże, s. 135.

18 J. Kalbach, *Projektowanie nawigacji strony WWW: optymalizacja funkcjonalności witryny*, Gliwice 2012, s. 147–155.

19 Tamże, s. 165–166; L. Rosenfeld, P. Morville, J. Arango, dz. cyt., s. 138–152.

20 L. Rosenfeld, P. Morville, J. Arango, dz. cyt., s. 125.

- Opcje systemu nawigacji reprezentują zawartość serwisu, grupy czy pojedyncze obiekty w systemach nawigacji i odgrywają kluczową rolę w podejmowaniu decyzji, który fragment przestrzeni informacyjnej zobaczy użytkownik. Etykiety te powinny być najbardziej przemyślane i spójne w całym systemie.
- Hasła skorowidza, które nie zawsze są widoczne dla użytkowników, ale odpowiadają za sprawne wyszukiwanie, przeglądanie i filtrowanie wyświetlanych zasobów. Są to m.in. metadane, tagi, słowa kluczowe.

Ze względu na charakter poszczególnych etykiet, generatywna sztuczna inteligencja ma szansę okazać się najbardziej pomocna w proponowaniu nagłówków, opcji systemów nawigacji i hasłach skorowidza. Aby się przekonać o możliwościach SI w tym zakresie, przeprowadzony zostanie eksperyment na kilku chatbotach.

Cel, problemy i metoda badawcza

Celem eksperymentu jest rozpoznanie możliwości generatywnej sztucznej inteligencji w zakresie tworzenia systemów etykietowania. Cel ten prowadzi do sformułowania następujących problemów badawczych:

- Czy generowane etykiety są zgodne z zasadami ich projektowania?
- Czy występują różnice w jakości etykiet tworzonych przez różne modele SI?
- Czy wygenerowane etykiety wymagają ingerencji człowieka w celu doprecyzowania ich znaczenia?

Aby odpowiedzieć na postawione pytania, przeanalizowane zostaną możliwości kilku wybranych chatbotów dla generowania etykiet w systemach nawigacji oraz nagłówków. Wygenerowane zestawy etykiet zostaną zebrane w tabelach, a następnie poddane analizie porównawczej za pomocą zestawu kryteriów. Kryteria oceny ustalone zostały na podstawie cech dobrych etykiet[21] oraz heurystyk i listy kontrolnej do oceny nawigacji internetowej, zaproponowanych przez Jamesa Kalbacha[22]. Ocenie poddany został także stopień spełnienia wymogów, które wystąpiły w promptach. Kryteria zostały wybrane z literatury źródłowej w taki sposób, aby można było ocenić etykiety pozostające wyłącznie w formie tekstowej i stanowiące wycinek systemu etykietowania. Do oceny zostały wybrane następujące warunki:

1) odpowiednia długość (etykieta składająca się z maksymalnie trzech słów);
2) spójność stylu i składni (stosowanie wielkich/małych liter, etykiety tworzone na jednej bazie np. na podstawie rzeczownika, czasownika lub inna konstrukcja);
3) utrzymanie poziomu szczegółowości (na danym poziomie podziału etykiety powinny być tak samo szczegółowe, jak treści dostarczone w poszczególnych eksperymentach);
4) wzajemne wykluczanie się etykiet (etykiety są jednoznaczne, a użytkownik nie ma wątpliwości, w której kategorii znajduje się poszukiwana informacja);
5) opisowość (etykiety są zrozumiałe i można przewidzieć, dokąd prowadzą odnośniki, jakie treści zostaną wyświetlone).

W każdym z kryteriów chatboty mogły otrzymać 0–2 punkty dla każdej z etykiet, maksymalnie 10 punktów dla jednej etykiety. Wartości liczbowe oznaczają:

21 J. Kalbach, dz. cyt.; L. Rosenfeld, P. Morville, J. Arango, dz. cyt.

22 J. Kalbach, dz. cyt.

2 – kryterium w pełni spełnione, 1 – kryterium spełnione częściowo, 0 – kryterium niespełnione.

Do eksperymentu wykorzystane zostały wyselekcjonowane chatboty. W przypadku projektowania systemów etykietowania ważna jest obsługa poszczególnych języków, w tym przypadku języka polskiego. Obecnie na rynku polskim udostępnione są m.in.: ChatGPT, Claude, Copilot, Gemini, Llama. Po wstępnych testach wymienionych chatbotów do eksperymentu wybrane zostały: Chat GPT w wersjach 4.o (bezpłatna, z dziennymi limitami) oraz 4.0 (płatna), Gemini w wersjach free (bezpłatna) i advanced (płatna) oraz Claude 3.5 Sonnet (bezpłatna, z dziennymi limitami). Copilot nie generował krótkich etykiet zgodnie z zadanym promptem, a Llama nie radził sobie z odpowiedziami w zadanym języku – na dane wejściowe w języku polskim odpowiadał w języku angielskim, dlatego nie zostały uwzględnione w eksperymencie.

Przebieg i opis eksperymentów

Każdy z chatbotów, niezależnie od testowanej wersji, otrzymał w eksperymentach taki sam prompt, w tym kontekst, zbiór danych wejściowych i wytyczne, co ma wygenerować na wyjściu. W promptach znalazły się opisy dotyczące samych etykiet: wskazano ich długość, podkreślono zachowanie spójności z całym systemem (tylko pod względem językowym, nie brano pod uwagę aspektu wizualnego), ich wzajemne wykluczanie się i zrozumiałość dla docelowych odbiorców. W poszczególnych eksperymentach chatboty otrzymały cztery grupy danych wejściowych. Każdy z promptów był wysłany tylko raz, bez dodatkowych poleceń modyfikowania wygenerowanej etykiety. Eksperymenty zostały przeprowadzone 9 i 10 lipca 2024 r.

Eksperyment 1
W pierwszym eksperymencie chatboty miały za zadanie wygenerować etykietę nadrzędną, czyli nazwę kategorii dla grupy powiązanych ze sobą elementów. Każdy z chatbotów otrzymał następujący prompt:

> Projektujesz system nawigacji dla serwisu internetowego uniwersytetu dotyczący współpracy międzynarodowej. Twoim zadaniem jest zaproponowanie etykiet tekstowych. Etykiety mają być: krótkie (maksymalnie trzy słowa), spójne z całym systemem (składnia i poziom szczegółowości), wzajemnie się wykluczające, zrozumiałe dla docelowych odbiorców (pracowników uczelni i studentów). Zaproponuj etykietę nadrzędną, która obejmie następującą grupę informacji (podstron):

Po symbolu „:" każdy chatbot otrzymał cztery kolejne zestawy danych wejściowych, na podstawie których zaproponował jedną etykietę nadrzędną.

Dane wejściowe 1.1:
- Wyjazdy zagraniczne (Studenci; Doktoranci; Nauczyciele akademiccy; Pozostali pracownicy);
- Przyjazdy osób z zagranicy (Goście zagraniczni; Profesorowie wizytujący; Umowy cywilno-prawne; Studenci zagraniczni);

– Erasmus+ (Studenci; Doktoranci; Nauczyciele akademiccy; Pozostali pracownicy).

Rezultat pracy chatbotów prezentuje tabela 1. W przypadku pierwszego zestawu danych chatboty zaproponowały tylko dwie etykiety: współpraca/mobilność międzynarodowa. Dla tego zestawu danych odpowiednie jest słowo mobilność, gdyż grupy informacji wejściowych dotyczą wyjazdów i programów zagranicznych oraz przyjazdów osób z zagranicy. ChatGPT 4.o i Gemini w obu wersjach wygenerowały też zestaw etykiet podrzędnych (powtórzyły większość danych wejściowych), co było czynnością zbędną.

Tabela 1. Wyniki eksperymentu 1 – generowania etykiet nadrzędnych – dla pierwszego zestawu danych

ChatGPT 4.o	ChatGPT Plus	Gemini	Gemini Advanced	Claude Sonnet
Współpraca Międzynarodowa	Mobilność międzynarodowa	Współpraca międzynarodowa	Współpraca międzynarodowa	Mobilność międzynarodowa

Źródło: opracowanie własne.

Dane wejściowe 1.2:
– Umowy o współpracy (Procedura zawierania umów; Lista aktualnych umów);
– Organizacje i stowarzyszenia międzynarodowe (Procedura przystąpienia; Lista aktualnych organizacji członkowskich);
– Akredytacje zagraniczne (Procedura ubiegania się o akredytacje; Lista aktualnych akredytacji);
– Podwójne dyplomy (Procedura podpisywania umów; Lista aktualnych umów).

Rezultat pracy chatbotów prezentuje tabela 2. Dla tego zestawu danych zaproponowane zostały trzy etykiety. Bezpłatne wersje ChatGPT i Claude wygenerowały „współpracę instytucjonalną", a płatny Chat GPT taką samą etykietę, co Gemini Free „Współpraca międzynarodowa". Z kolei etykieta Gemini Advanced jest zbyt ogólna. Warto zwrócić uwagę, że poszczególne chatboty generują etykiety zapisywane różną wielkością liter.

Tabela 2. Wyniki eksperymentu 1 – generowania etykiet nadrzędnych – dla drugiego zestawu danych

ChatGPT 4.o	ChatGPT Plus	Gemini	Gemini Advanced	Claude Sonnet
Współpraca Instytucjonalna	Współpraca międzynarodowa	Współpraca międzynarodowa	Międzynarodowe	Współpraca instytucjonalna

Źródło: opracowanie własne.

Dane wejściowe 1.3:
– Informacje o uczelni;
– Udział uczelni w targach;
– Uczelnia w rankingach międzynarodowych.

Rezultat pracy chatbotów prezentuje tabela 3. Pomimo że zestaw danych wejściowych był niewielki, chatboty wykazały się dużą kreatywnością, a ich propozycje są zróżnicowane. Taką samą etykietę wygenerowały ChatGPT Plus oraz Claude. Gemini tym razem zignorował zadaną długość etykiet, ale jego etykieta dobrze oddaje

zawartość kategorii. Z kolei etykieta Gemini Advanced po raz kolejny okazała się najbardziej ogólna, a przez to mało opisowa.

Tabela 3. Wyniki eksperymentu 1 – generowania etykiet nadrzędnych – dla trzeciego zestawu danych

ChatGPT 4.o	ChatGPT Plus	Gemini	Gemini Advanced	Claude Sonnet
Promocja Uczelni	Profil międzynarodowy	Uniwersytet na arenie międzynarodowej	O uczelni	Profil międzynarodowy

Źródło: opracowanie własne.

Dane wejściowe 1.4:
- Wyjazdy na praktykę (SMP);
- Wyjazdy na studia (SMS);
- Wyjazdy w ramach Programu Edukacja;
- Zawieranie umów międzyuczelnianych Erasmus+ Programme Countries.

Rezultat pracy chatbotów prezentuje tabela 4. Ostatni pakiet danych wejściowych w eksperymencie dotyczył programu Erasmus+ dla studentów. Jedynie ChatGPT Plus odwzorował taką zawartość w zaproponowanej etykiecie. Z kolei trzy chatboty wygenerowały etykietę „Mobilność międzynarodowa", co jednak jest szerszym pojęciem i stanowi najlepszą etykietę dla pierwszego zestawu danych.

Tabela 4. Wyniki eksperymentu 1 – generowania etykiet nadrzędnych – dla czwartego zestawu danych

ChatGPT 4.o	ChatGPT Plus	Gemini	Gemini Advanced	Claude Sonnet
Mobilność Międzynarodowa	Programy Erasmus+	Mobilność międzynarodowa	Mobilność Międzynarodowa	Programy wymian

Źródło: opracowanie własne.

Eksperyment 2

W drugim eksperymencie chatboty miały za zadanie wygenerować etykietę funkcjonującą jako nagłówek dla zadanego tekstu. Teksty dotyczyły współpracy międzynarodowej i zostały zaczerpnięte z serwisu internetowego Uniwersytetu Śląskiego (https://us.edu.pl), jednak występująca w opisach nazwa własna uczelni została zamieniona na bardziej ogólne hasła typu 'uczelnia' lub 'uniwersytet'. Każdy z chatbotów otrzymał następujący prompt:

> Projektujesz system nawigacji dla serwisu internetowego uniwersytetu dotyczący współpracy międzynarodowej. Twoim zadaniem jest zaproponowanie etykiet tekstowych. Etykiety mają być: krótkie (maksymalnie trzy słowa), spójne z całym systemem (składnia i poziom szczegółowości), wzajemnie się wykluczające, zrozumiałe dla docelowych odbiorców (pracowników uczelni i studentów). Zaproponuj nagłówek dla poniższego tekstu, który będzie jednocześnie pełnił funkcję etykiety w systemie nawigacji:

Ponownie po symbolu „:" każdy chatbot otrzymał cztery kolejne zestawy danych wejściowych, na podstawie których zaproponował jedną etykietę.

Dane wejściowe 2.1:

> Międzynarodowe umowy o współpracy odzwierciedlają charakter działalności uczelni. Są to porozumienia zawierane w celu realizacji międzynarodowych grantów i projektów badawczych, badań naukowych, programów mobilnościowych, jak również prowadzenia wspólnych studiów czy podejmowania działań propagujących język polski poza granicami kraju.
> Wśród partnerów uniwersytetu znajdują się uczelnie z całego świata.
> W ramach programu Erasmus+ uczelnia zawarła ponad 750 porozumień, które umożliwiają wyjazdy i przyjazdy studentów, doktorantów, pracowników naukowych i kadry administracyjnej.

Rezultat pracy chatbotów prezentuje tabela 5. Trzy chatboty wygenerowały taką samą etykietę – „Umowy międzynarodowe". Taka etykieta odpowiada dostarczonej zawartości. Pozostałe dwie propozycje również dobrze opisują zawartość, jednak etykieta wygenerowana przez Gemini jest zbyt długa. Nagłówek, który obecny jest na stronie źródłowej to: Umowy bilateralne. Taka etykieta, mimo iż jest ogólnie przyjętym terminem, może być niezrozumiała dla niektórych odbiorców.

Tabela 5. Wyniki eksperymentu 2 – generowania nagłówków – dla pierwszego zestawu danych

ChatGPT 4.o	ChatGPT Plus	Gemini	Gemini Advanced	Claude Sonnet
Międzynarodowe Umowy	Umowy międzynarodowe	Międzynarodowe umowy o współpracy	Umowy międzynarodowe	Umowy międzynarodowe

Źródło: opracowanie własne.

Dane wejściowe 2.2:

> Współpraca i wymiana międzynarodowa to ważne aspekty działalności akademickiej. Dziś trudno wyobrazić sobie uniwersytet bez studentów i pracowników z zagranicy, programów w językach obcych, mobilności zagranicznej studentów i kadry akademickiej oraz międzynarodowych projektów i badań. Prowadzona przez uczelnię intensywna współpraca międzynarodowa pozwala na wymianę akademicką oraz podejmowanie różnorodnych przedsięwzięć edukacyjnych i naukowych wspólnie z partnerami z ponad 90 krajów świata. Corocznie z oferty wymiany akademickiej korzysta około 700 studentów i doktorantów, a także około 2000 pracowników, których wyjazdy związane są z realizacją badań naukowych, projektów międzynarodowych, udziałem w konferencjach, a także działalnością edukacyjną i dydaktyczną. Jednocześnie naszą uczelnię odwiedza około 2800 pracowników naukowych, dydaktycznych i administracyjnych reprezentujących instytucje partnerskie. Programy i projekty mobilnościowe.

Załączony fragment dotyczył ogólnych informacji o wymianie międzynarodowej uczelni. Rezultat pracy chatbotów prezentuje tabela 6. Wygenerowane nagłówki różnią się od siebie, ale poza propozycjami ChatGPT 4.o (zbyt ogólna etykieta) i Claude (etykieta nie informuje o wymianie) dobrze oddają zawartość załączonego tekstu. Gemini po raz kolejny wygenerował dłuższą etykietę. Autorski nagłówek brzmi „Mobilność międzynarodowa", jednak inne nieopublikowane badania prowadzone przez autorkę, wykazały, że słowo mobilność nie jest właściwie

rozumiane przez osoby studiujące. Wygenerowane nagłówki powinny być lepiej rozumiane przez część docelowej grupy, dlatego propozycje chatbotów mogą sprawdzić się lepiej.

Tabela 6. Wyniki eksperymentu 2 – generowania nagłówków – dla drugiego zestawu danych

ChatGPT 4.o	ChatGPT Plus	Gemini	Gemini Advanced	Claude Sonnet
Współpraca i Wymiana	Wymiana akademicka	Współpraca i wymiana międzynarodowa	Wymiana międzynarodowa	Współpraca międzynarodowa

Źródło: opracowanie własne.

Dane wejściowe 2.3:

> Każdy doktorant w Szkole Doktorskiej ma możliwość rozszerzenia swojego kształcenia i prowadzenia swoich badań w międzynarodowym środowisku akademickim i naukowym. W zależności od potrzeb, może to być udział w zajęciach akademickich, inne formy kształcenia i wymiany wiedzy, staż, czy przeprowadzenie części lub całości swoich badań we współpracy z instytucjami zagranicznymi. Istotnym aspektem jest także prezentacja swoich osiągnięć na międzynarodowych i zagranicznych konferencjach naukowych. W przypadku, kiedy całość projektu doktorskiego powstaje we współpracy z innym ośrodkiem akademickim, sam doktorat może mieć charakter międzynarodowy, to znaczy może być to stopień nadany równocześnie przez Uniwersytet i drugi ośrodek. Obecnie Szkoła Doktorska jest na etapie nawiązywania kontaktów z partnerami zagranicznymi.
> Rodzaje aktywności międzynarodowej.

Tym razem załączony tekst dotyczył możliwości prowadzenia badań przez doktorantów i ich kształcenia w skali międzynarodowej. Rezultat pracy chatbotów prezentuje tabela 7. Najlepiej treść oddaje nagłówek wygenerowany przez Gemini, jednak nie spełnia on wytycznej dotyczącej długości etykiety. Pod względem opisowości dobrze wypadł też nagłówek zaproponowany przez ChatGPT 4.o. Nagłówek „Doktoranci za granicą" częściowo charakteryzuje dane wejściowe, jednak Claude Sonnet zaproponował etykietę z błędem – łącznym zapisem słów „zagranicą". Etykieta zaproponowana przez ChatGPT Plus może sugerować zupełnie inną zawartość – pewien rodzaj doktoratu o międzynarodowym charakterze. Oryginalna etykieta w źródle to „Działania międzynarodowe", jednak wygenerowane nagłówki precyzyjniej wskazywały, że w tekście chodzi o doktorantów.

Tabela 7. Wyniki eksperymentu 2 – generowania nagłówków – dla trzeciego zestawu danych

ChatGPT 4.o	ChatGPT Plus	Gemini	Gemini Advanced	Claude Sonnet
Aktywność Międzynarodowa Doktorantów	Doktorat międzynarodowy	Międzynarodowe możliwości dla doktorantów	Doktoranci za granicą	Doktoranci zagranicą

Źródło: opracowanie własne.

Dane wejściowe 2.4:

> Uniwersytet jest członkiem wielu międzynarodowych organizacji akademickich. Przedstawiciele uczelni zasiadają w komitetach sterujących i pełnią funkcję ekspertów w sieciach i stowarzyszeniach zrzeszających społeczność akademicką z całego świata. Uczelnia współpracuje z kluczowymi organizacjami, wśród których warto wymienić: European University Association (EUA), L'Agence universitaire de la Francophonie (AUF) czy European University Information System (EUNIS). Uczelnia należy do takich organizacji, jak: AUF – Agence Universitaire de la Francophonie; CIOFF – Conseil International des Organisations de Festivals de Folklore et d'Art Traditionnels; CIRCEOS – Interuniversity Centre for Research and Cooperation with Eastern and South-Eastern Europe; EASSW – European Association of Schools of Social Work; ECPR – European Consortium for Political Research; ELFA – European Law Faculties Association; EUA – European University Association; EUCUNET – European Universities Network; EUNIS – European University Information Systems; EUPRIO – European Association of Communication Professionals in Higher Education; EUSEA – European Science Engagement Association; GSA – Genetics Society of AustralAsia Inc.; IAAP – International Association of Applied Psychology; World Piano Teachers Association. Uniwersytet jest również członkiem Konsorcjum Śląskich Uczelni Publicznych (SUN – Silesian Universities Network), działającego na rzecz promocji szkół wyższych w kraju, regionie oraz na arenie międzynarodowej. W ramach krajowej sieci IROs Forum (International Relations Offices Forum) uczestniczy w inicjatywach, których celem jest zwiększenie jakości, efektywności i zakresu współpracy międzynarodowej polskich uczelni.

Ostatni z tekstów to przede wszystkim wykaz organizacji międzynarodowych, z którymi współpracuje uczelnia. Rezultat pracy chatbotów prezentuje tabela 8. Pomimo czterech różnych propozycji nagłówków, każdy w nieco inny sposób dobrze oddaje zawartość załączonego tekstu. Gemini po raz kolejny zaproponował dłuższą etykietę niż zakładano. Etykieta występująca na stronie źródłowej to: „Organizacje członkowskie".

Tabela 8. Wyniki eksperymentu 2 – generowania nagłówków – dla czwartego zestawu danych

ChatGPT 4.o	ChatGPT Plus	Gemini	Gemini Advanced	Claude Sonnet
Członkostwo w Organizacjach	Sieci międzynarodowe	Międzynarodowe organizacje i stowarzyszenia	Organizacje międzynarodowe	Sieci międzynarodowe

Źródło: opracowanie własne.

Samo zestawienie rezultatów w tabelach nie pozwala w pełni ocenić, jak chatboty radzą sobie z generowaniem etykiet. Różnie zinterpretowały poszczególne zestawy danych, dla niektórych proponowały bardzo podobne etykiety, przy innych zestawach trudno znaleźć podobieństwa. ChatGPT 4.o i Gemini w obu wersjach wygenerowały też zestawy etykiet podrzędnych (powtarzając większość danych wejściowych), co stanowiło niepotrzebną czynność. Poza ChatGPT 4.o, pozostałe chatboty mniej lub bardziej szczegółowo uzasadniały swoje odpowiedzi. Aby ocenić, jak sprawdzają się poszczególne chatboty w generowaniu obu rodzajów etykiet, przeprowadzona zostanie analiza porównawcza.

Analiza wyników

Oceny wygenerowanych etykiet dokonano poprzez analizę porównawczą. Zestawy etykiet zostały zebrane w dwóch tabelach – oddzielnie dla każdego eksperymentu – a następnie poddane ocenie punktowej w skali 0–2. Kryteria poddawane ocenie to kolejno:

1) długość etykiet;
2) spójność stylu i składni;
3) utrzymanie poziomu szczegółowości;
4) wzajemne wykluczanie się etykiet;
5) opisowość.

Tabela 9. Porównanie wyników eksperymentu 1 – generowania etykiet nadrzędnych – punkty dla poszczególnych etykiet

	ChatGPT 4.o	**ChatGPT Plus**	**Gemini**	**Gemini Advanced**	**Claude Sonnet**
Etykieta	Współpraca Międzynarodowa	Mobilność międzynarodowa	Współpraca międzynarodowa	Współpraca międzynarodowa	Mobilność międzynarodowa
Punkty	2/2/2/1/0	2/2/2/2/2	2/2/2/0/0	2/1/2/2/0	2/2/2/2/2
Etykieta	Współpraca Instytucjonalna	Współpraca międzynarodowa	Współpraca międzynarodowa	Międzynarodowe	Współpraca instytucjonalna
Punkty	2/2/2/1/1	2/2/2/2/2	2/2/2/0/2	2/0/1/1/0	2/2/2/2/1
Etykieta	Promocja Uczelni	Profil międzynarodowy	Uniwersytet na arenie międzynarodowej	O uczelni	Profil międzynarodowy
Punkty	2/2/2/2/2	2/2/2/2/1	0/1/2/2/1	2/0/1/2/1	2/1/2/2/1
Etykieta	Mobilność Międzynarodowa	Programy Erasmus+	Mobilność międzynarodowa	Mobilność międzynarodowa	Programy wymian
Punkty	2/2/2/2/1	2/1/2/2/2	2/2/2/2/1	2/1/2/2/1	2/1/2/2/1
Suma punktów	**34**	**38**	**29**	**25**	**35**

Źródło: opracowanie własne.

W generowaniu etykiet nadrzędnych, zbiorczo zestawionych w tabeli 9, najlepiej wypadł ChatGPT Plus (38/40 punktów), uzyskując maksymalną liczbę punktów w trzech kryteriach (długość etykiet, szczegółowość i wykluczanie się) dla wszystkich etykiet. W propozycjach etykiet dominuje zapis pierwszego wyrazu wielką literą, a pozostałych małymi, z tej formy wyłamał się jedynie ChatGPT 4.o. W ogólnym zestawieniu dobrze poradziły sobie również bezpłatne wersje Claude (35 p.) i ChatGPT (34 p.). Claude miał problemy w zachowaniu spójności składni i opisowości. Z kolei niektóre etykiety wygenerowane przez ChatGPT 4.o nie wykluczały się w pełni, powstał też problem z ich opisowością. Najsłabiej w ocenie wypadł zestaw etykiet zaproponowany przez Gemini Advanced (25/40). Wystąpiły w nich problemy ze spójnością, utrzymaniem tego samego poziomu szczegółowości i opisowością. Nieco lepiej poradził sobie Gemini w wersji bezpłatnej, tracąc punkty dla każdej z etykiet w różnych kategoriach. W eksperymencie niemal we wszystkich przypadkach chatboty wygenerowały etykiety o odpowiedniej długości i na tym samym poziomie szczegółowości (tabela 10). Jedynie Gemini Advanced

miał problem z dostosowaniem się do tych wytycznych. Z kolei najsłabiej chatboty wypadły w zakresie opisowości, czyli odzwierciedlenia zawartości. Jednak trzeba pamiętać, że jest to najbardziej subiektywne kryterium, a znaczenie jest różnie interpretowane przez użytkowników.

Tabela 10. Porównanie wyników eksperymentu 1 – generowania etykiet nadrzędnych – punkty dla poszczególnych kryteriów

	ChatGPT 4.o	**ChatGPT Plus**	**Gemini**	**Gemini Advanced**	**Claude Sonnet**	**Suma punktów**
długość etykiet	2/2/2/2	2/2/2/2	2/2/0/2	2/2/2/2	2/2/2/2	38
spójność stylu i składni	2/2/2/2	2/2/2/1	2/2/1/2	1/0/0/1	2/2/1/1	30
utrzymanie poziomu szczegółowości	2/2/2/2	2/2/2/2	2/2/2/2	2/1/1/2	2/2/2/2	38
wzajemne wykluczanie się etykiet	1/1/2/2	2/2/2/2	0/0/2/2	2/1/2/2	2/2/2/2	33
opisowość	0/1/2/1	2/2/1/2	0/2/1/1	0/0/1/1	2/1/1/1	22

Źródło: opracowanie własne.

Nieco inaczej wygląda sytuacja dla etykiet generowanych jako nagłówki dla krótkich tekstów (tabela 11). Dla pierwszego tekstu wygenerowane etykiety były identyczne w trzech przypadkach i dla nagłówka „Umowy międzynarodowe" chatboty otrzymały maksymalną liczbę punktów. Dwie pozostałe etykiety w ocenie punktowej też wypadły dobrze.

Tabela 11. Porównanie wyników eksperymentu 2 – generowania nagłówków nadrzędnych – punkty dla poszczególnych etykiet

	ChatGPT 4.o	**ChatGPT Plus**	**Gemini**	**Gemini Advanced**	**Claude Sonnet**
Etykieta	Międzynarodowe Umowy	Umowy międzynarodowe	Międzynarodowe umowy o współpracy	Umowy międzynarodowe	Umowy międzynarodowe
Punkty	2/1/2/2/2	2/2/2/2/2	0/2/2/2/2	2/2/2/2/2	2/2/2/2/2
Etykieta	Współpraca i Wymiana	Wymiana Akademicka	Współpraca i wymiana międzynarodowa	Wymiana międzynarodowa	Współpraca międzynarodowa
Punkty	2/1/1/2/1	2/1/2/2/2	0/1/1/2/2	2/1/2/2/2	2/1/2/2/1
Etykieta	Aktywność Międzynarodowa Doktorantów	Doktorat międzynarodowy	Międzynarodowe możliwości dla doktorantów	Doktoranci za granicą	Doktoranci zagranicą
Punkty	2/2/2/2/2	2/1/2/2/0	0/2/2/2/2	2/1/2/2/1	2/1/2/2/1
Etykieta	Członkostwo w Organizacjach	Sieci międzynarodowe	Międzynarodowe organizacje i stowarzyszenia	Organizacje międzynarodowe	Sieci międzynarodowe
Punkty	2/2/2/2/2	2/2/2/2/1	0/2/2/2/1	2/2/2/2/1	2/1/2/2/1
Suma punktów	**36**	**35**	**29**	**36**	**34**

Źródło: opracowanie własne.

W zadaniu najlepsze okazały się ChatGPT w wersji 4.o (36/40) oraz Gemini Advanced (36/40). W przypadku pierwszego narzędzia dwie etykiety uzyskały maksymalną liczbę punktów, a najgorzej wypadła druga, zbyt ogólna etykieta. Z kolei etykiety Gemini Advanced miały usterki w zakresie spójności i opisowości. ChatGPT Plus zyskał tylko jeden punkt mniej (35/40), tracąc w zakresie spójności i opisowości, a najmniej punktów otrzymała etykieta „Doktorat międzynarodowy", która może sugerować pewien rodzaj doktoratu, podobnie jak matura międzynarodowa. Dobrze ocenione zostały etykiety Claude Sonnet (34/40), a problemy ponownie pojawiły się w spójności i opisowości. Jedna etykieta zawiera błąd językowy – zapis łączny „zagranicą".

Tabela 12. Porównanie wyników eksperymentu 2 – generowania nagłówków nadrzędnych – punkty dla poszczególnych kryteriów

	ChatGPT 4.o	ChatGPT Plus	Gemini	Gemini Advanced	Claude Sonnet	Suma punktów
długość etykiet	2/2/2/2	2/2/2/2	0/0/0/0	2/2/2/2	2/2/2/2	32
spójność stylu i składni	1/1/2/2	2/1/1/2	2/1/2/2	2/1/1/2	2/1/1/1	30
utrzymanie poziomu szczegółowości	2/1/2/2	2/2/2/2	2/1/2/2	2/2/2/2	2/2/2/2	38
wzajemne wykluczanie się etykiet	2/2/2/2	2/2/2/2	2/2/2/2	2/2/2/2	2/2/2/2	40
opisowość	2/1/2/2	2/2/0/1	2/2/2/1	2/2/1/1	2/1/1/1	30

Źródło: opracowanie własne.

Z kolei dla poszczególnych kryteriów oceny (tabela 12), generowane nagłówki okazały się najlepsze pod względem wzajemnego wykluczania. Świetny wynik został osiągnięty w zakresie utrzymania poziomu szczegółowości. W całym zadaniu Gemini Free stracił punkty w pierwszym kryterium, gdyż etykiety były dłuższe niż w prompcie. Długość etykiet została spełniona w pełni dla czterech narzędzi. Najsłabiej chatboty wypadły ponownie w zakresie opisowości, żaden z nich nie poradził sobie całkowicie ze spójnością stylu i składni.

Wnioski

Analizując wyniki punktowe uzyskane w pięciu kategoriach można odpowiedzieć na wszystkie problemy badawcze. Po pierwsze testowane chatboty w różnym stopniu stosują się do zasad projektowania, mimo że zostały one zawarte w promptach, a wyniki zmieniają się dla poszczególnych zestawów danych wejściowych. Niezależnie od eksperymentu, chatboty najlepiej poradziły sobie z długością etykiet i, poza Gemini Free, spełniały to kryterium w 100%. Zestawy danych wejściowych pozwalały na generowanie etykiet nadrzędnych na tym samym poziomie szczegółowości. Z kolei fragmenty tekstów zapewniły wzajemne wykluczanie się etykiet wygenerowanych w drugim eksperymencie. Niezależnie od danych wejściowych, największym problemem okazała się opisowość etykiet oraz spójność stylu i składni, co daje odpowiedź na trzeci problem badawczy – wygenerowane etykiety wymagają korekt człowieka.

Biorąc pod uwagę drugi problem badawczy, można zauważyć różnice w jakości generowanych etykiet przez poszczególne narzędzia. Sumarycznie najlepiej

w generowaniu etykiet obu rodzajów poradził sobie ChatGPT Plus (73/80 p.), tracąc punkty w drugim eksperymencie w aspekcie spójności składniowej i opisowości. Niewiele mniej punktów otrzymały bezpłatne narzędzia: ChatGPT 4.o (70 p.) i Claude Sonnet (69 p.), tracąc punkty w tych samych kategoriach, co płatna wersja ChataGPT. Najdłuższe, a przez to niezgodne z założeniami etykiety generował Gemini Free, w zestawach etykiet pojawiały się też problemy ze spójnością i opisowością. Z kolei Gemini Advanced nie poradził sobie z etykietami nadrzędnymi w eksperymencie pierwszym, ale do realizacji drugiego zadania okazał się dobrym narzędziem.

Analizując wyniki i wyciągając wnioski, należy pamiętać o ograniczeniach przeprowadzonych badań. Eksperymenty były przeprowadzone na małej liczbie zestawów danych (po cztery zestawy). Etykiety zaproponowane przez chatboty nie były edytowane – brana pod uwagę była wyłącznie pierwsza wersja, a wszelkie odpowiedzi można dopracowywać w kolejnych dyskusjach. Przeprowadzona analiza porównawcza bazowała na kilku wybranych kryteriach, częściowo jest to ocena subiektywna (np. opisowość przez inne osoby mogłaby zostać oceniona inaczej). Być może inny zestaw danych wejściowych, czy kryteriów do oceny dałyby odmienne rezultaty dla poszczególnych chatbotów.

Podsumowując, na podstawie tego niewielkiego materiału badawczego większość chatbotów okazała się dobrym narzędziem do generowania etykiet i może stanowić źródło inspiracji w projektowaniu systemów etykietowania. ChatGPT w obu wersjach oraz Claude Sonnet wypadły dobrze w obu eksperymentach, uzyskując w ocenie co najmniej 86% punktów. Gemini Advanced okazał się bardzo dobrym narzędziem w generowaniu etykiet w formie nagłówków. Analizując uzyskane punkty i poszczególne odpowiedzi, można uznać, że jako źródło inspiracji architekta informacji wystarczą bezpłatne wersje chatbotów. Najlepiej skorzystać z dwóch narzędzi i porównać otrzymane rezultaty, a następnie wybrać najlepszą wersję. Generowane zestawy etykiet mogą być dopracowywane w ramach kolejnych poleceń wysyłanych chatbotom. Generatywna sztuczna inteligencja cały czas się uczy i rozwija, więc poszczególne aplikacje z czasem mogą stawać się coraz lepszymi partnerami w projektowaniu systemów etykietowania. Warto nadmienić, że chatboty mogą być szczególnie przydatne w przypadku projektowania serwisów wielojęzycznych.

Niezależnie od stosowanych źródeł inspiracji, należy pamiętać, że nie można przejmować wygenerowanych etykiet bez zastanowienia. Osoba projektująca system etykietowania powinna znać zawartość przestrzeni informacyjnej, stosowaną terminologię i docelowych użytkowników, by do nich dostosować słownictwo. Niezastąpione są badania z użytkownikami, bo to oni ostatecznie oceniają zrozumiałość i skuteczność systemu etykietowania.

Bibliografia

AI Tools in Framer, [on-line:] https://www.framer.com/features/ai – 16.07.2024.

Akhtar Z.B., *Unveiling the evolution of generative AI (GAI): a comprehensive and investigative analysis toward LLM models (2021–2024) and beyond*, „Journal of Electrical Systems and Information Technology" 2024, vol. 11, no. 1, https://doi.org/10.1186/s43067-024-00145-1.

Carr D., *ChatGPT tops 25 million daily visits*, Similarweb, 3.02.2023, [on-line:] https://www.similarweb.com/blog/insights/ai-news/chatgpt-25-million – 11.07.2024.

Co to jest duży model języka (LLM)?, OVHcloud Polska, [on-line:] https://www.ovhcloud.com/pl/learn/what-is-large-language-model – 7.07.2024.

Dwivedi Y.K., Kshetri N., Hughes L., i in., *Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy*, „International Journal of Information Management" 2023, vol. 71, https://doi.org/10.1016/j.ijinfomgt.2023.102642.

FigJam AI, [on-line:] https://www.figma.com/figjam/ai – 16.07.2024.

Giray L., *Prompt engineering with ChatGPT: A guide for academic writers*, „Annals of Biomedical Engineering" 2023, vol. 51, no. 12, s. 2629–2633, https://doi.org/10.1007/s10439-023-03272-4.

Kalbach J., *Projektowanie nawigacji strony WWW: optymalizacja funkcjonalności witryny*, Gliwice 2012.

Meet Hotjar AI: Your personal AI-powered research assistant, [on-line:] https://www.hotjar.com/blog/introducing-hotjar-ai – 16.07.2024.

Nielsen J., *What AI can and cannot do for UX*, 1.02.2024, [on-line:] https://www.uxtigers.com/post/ai-can-cannot-do-ux – 4.06.2024.

Pirker M., *AI Insights: Instantly uncover key insights and patterns across your user tests*, 7.06.2024, [on-line:] https://www.userbrain.com/blog/ai-insights-instantly-uncover-key-insights-and-patterns-across-your-user-tests – 16.07.2024.

Rosenfeld L., Morville P., Arango J., *Architektura informacji w serwisach internetowych i nie tylko*, tłum. R. Meryk, Gliwice 2017.

Rustambek M., *The role of artificial intelligence in web sites*, „Proceedings of International Conference on Scientific Research in Natural and Social Sciences" 2023, vol. 2, no. 7, s. 101–107, [on-line:] https://econferenceseries.com/index.php/srnss/article/view/2407 – 9.07.2024.

Tomaszczyk J., Matysek A., *IA, UX, UID, IxD – analiza terminów i pojęć*, „Zagadnienia Informacji Naukowej – Studia Informacyjne" 2020, nr 1, s. 121–143.

Streszczenie

Celem artykułu jest zbadanie możliwości wykorzystania sztucznej inteligencji (SI) do automatycznego generowania etykiet w formie nagłówków i opcji systemów etykietowania.
Metody badań: Przeprowadzono dwa eksperymenty, w którym pięć chatbotów – ChatGPT 4.o, ChatGPT Plus, Gemini Free, Gemini Advanced, Claude Sonnet – miało za zadanie wygenerować zestawy etykiet zgodnie z zadanym promptem i czterema zestawami danych wejściowych. Ocenę możliwości chatbotów przeprowadzono stosując metodę analizy porównawczej. Analiza obejmowała ocenę długości, spójności, szczegółowości, wzajemnego wykluczania się i opisowości wygenerowanych etykiet.
Wyniki: Chatboty, w zależności od danych wejściowych, generowały podobne lub bardzo różne etykiety. W zakresie generowania opcji systemów nawigacji uzyskały następujące liczby punktów na 40 możliwych: ChatGPT 4.o – 34 p., ChatGPT Plus – 38 p., Gemini Free – 29 p., Gemini Advanced – 25 p., Claude Sonnet – 35 p. Nieco inaczej poradziły sobie w generowaniu nagłówków, otrzymując: ChatGPT 4.o – 36 p., ChatGPT Plus – 35 p., Gemini Free – 29 p., Gemini Advanced – 36 p., Claude Sonnet – 34 p. Sumarycznie najlepiej poradził sobie w eksperymentach ChatGPT Plus, a najsłabiej Gemini Free.

Wnioski: W generowaniu etykiet dobrze radzą sobie zarówno płatne, jak i bezpłatne wersje chatbotów, ale występują różnice w jakości rezultatów dla poszczególnych zestawów danych wejściowych. Niektóre z zasad projektowania chatboty spełniają w pełni, a z innymi sobie nie radzą. Każdy z testowanych chatbotów może być źródłem inspiracji architekta informacji, propozycje etykiet wymagają korekty ze strony człowieka, a ostateczne decyzje projektowe powinny zostać podjęte po przeprowadzeniu odpowiednich testów z użytkownikami.

Słowa kluczowe: architektura informacji, generowanie etykiet, systemy etykietowania, sztuczna inteligencja

The application of artificial intelligence in the automated design of labeling systems: An analysis of experiment results

Abstract

Thesis/aim: The paper aims to investigate the potential of using artificial intelligence (AI) to automatically generate labels in the form of headers and options for labeling systems.
Research methods: Two experiments were conducted, in which five chatbots – ChatGPT 4.0, ChatGPT Plus, Gemini Free, Gemini Advanced, and Claude Sonnet – were asked to generate sets of labels according to a given prompt and four sets of input data. The capabilities of chatbots were carried out using a comparative analysis method. The analysis included assessing the length, consistency, specificity, mutual exclusivity, and descriptiveness of the generated labels.
Results: The chatbots generated similar or very different labels depending on the input data. In terms of generating navigation system options, they scored the following points out of 40 possible: ChatGPT 4.0 – 34 pts, ChatGPT Plus – 38 pts, Gemini Free – 29 pts, Gemini Advanced – 25 pts, Claude Sonnet – 35 pts. They performed slightly differently in generating headings, receiving ChatGPT 4.0 – 36 pts, ChatGPT Plus – 35 pts, Gemini Free – 29 pts, Gemini Advanced – 36 pts, and Claude Sonnet – 34 pts. ChatGPT Plus performed best in the experiments, while Gemini Free performed the weakest.
Conclusions: Both paid and free versions of chatbots perform well in generating labels, but there are differences in the quality of results for different input datasets. They fully meet some design principles while struggling with others. Each of the tested chatbots can be a source of inspiration for an information architect, though the proposed labels require human review and correction, and final design decisions should be made after conducting appropriate user tests.

Keywords: information architecture, generating labels, labelling systems, artificial intelligence