Rozwiąż swój problem z przymrużeniem oka

AI podało źródła. Jedno nawet istnieje

Jak sprawdzać cytaty i źródła tworzone przez narzędzia AI



Max Paradox

AI podało źródła. Jedno nawet istnieje

Jak sprawdzać cytaty i źródła tworzone przez narzędzia AI

Max Paradox

PROBLEM? WYLUZUJ. OGARNIESZ TO!

ISBN: 9788384906248

ALL-IN Rafał Płonkowski



SPIS TREŚCI

INTRO 5

Rozdział 1 - Bibliografia z krainy fantazji, czyli dlaczego AI wymyśla źródła 10

Rozdział 2 - Cytat w cudzysłowie, prawda na urlopie 19

Rozdział 3 - Link działa, ale dokąd prowadzi? Anatomia fałszywego dowodu 28

Rozdział 4 - „Przecież AI to sprawdziła", czyli najdroższe wymówki świata 37

Rozdział 5 - Nie szukaj igły w bibliotece, czyli jak znaleźć źródło bez utraty zmysłów 48

Rozdział 6 - Kiedy odpowiedź pasuje aż za dobrze, czyli pułapka własnych oczekiwań 58

Rozdział 7 - Głuchy telefon w garniturze, czyli jak źródła giną między ludźmi i dokumentami 68

Rozdział 8 - PDF wszystko przyjmie, czyli kiedy AI źle czyta prawdziwe dokumenty 78

Rozdział 9 - Pięć minut dla prawdy, czyli jak przeprowadzić pierwszy audyt odpowiedzi AI 89

Rozdział 10 - Od cytatu do oryginału, czyli śledztwo z prawdziwym zakończeniem 99

Rozdział 11 - Liczby nie kłamią, ale ich opisy mają bujną wyobraźnię 109

Rozdział 12 - Zbuduj bibliografię, której nie trzeba się wstydzić 119

Rozdział 13 - Naucz AI pokazywać dowody, a nie robić dobre wrażenie 130

Rozdział 14 - Od odpowiedzi AI do decyzji, której da się bronić 140

Rozdział 15 - Dwa źródła, trzy prawdy i jeden człowiek bliski rozpaczy 152

Rozdział 16 - Źródło, które zaczyna wydawać rozkazy, czyli nie ufaj każdej instrukcji w dokumencie 162

Rozdział 17 - Kiedy nie da się sprawdzić wszystkiego, czyli sztuka uczciwego niedosytu 173

Rozdział 18 - Kontrola, która nie kończy się po kliknięciu „Wyślij” 184

Rozdział 19 - Własny radar na bzdury, czyli jak nie wrócić do starych nawyków 196

Rozdział 20 - AI może pisać pewnie. Ty możesz wiedzieć, skąd to wie 207

INTRO

Wyobraź sobie, że przygotowujesz ważny raport. Termin zbliża się z prędkością kuriera, który akurat przyjechał wtedy, gdy wszedłeś pod prysznic. Masz dane, masz argumenty, brakuje ci tylko kilku solidnych źródeł. Uruchamiasz narzędzie AI i prosisz o publikacje potwierdzające określone twierdzenie. Po chwili otrzymujesz pięć eleganckich pozycji bibliograficznych. Autorzy mają nazwiska, artykuły mają tytuły, czasopisma mają numery, a niektóre publikacje nawet identyfikatory DOI. Wszystko wygląda tak profesjonalnie, że najchętniej wydrukowałbyś odpowiedź, oprawił ją w skórę i postawił na półce obok encyklopedii.

Jest tylko drobny problem. Pierwszy artykuł nie istnieje. Drugi istnieje, ale dotyczy czegoś innego. Trzeci ma prawdziwego autora i wymyślony tytuł. Czwarty prowadzi do strony, która już dawno przestała działać. Piąty rzeczywiście istnieje, lecz jego autorzy twierdzą coś niemal przeciwnego do tego, co napisała AI.

Gratulacje. Właśnie otrzymałeś bibliografię, w której rzeczywistość występuje gościnnie.

Najbardziej niepokojące nie jest jednak to, że sztuczna inteligencja potrafi wymyślić źródło. Znacznie ciekawsze jest to, jak przekonująco potrafi je przedstawić. Nie pojawia się ostrzeżenie: „Uwaga, poniższe nazwiska mogą być rezultatem literackiej fantazji procesora". Nie słyszysz nerwowego kaszlnięcia. Nie widzisz cyfrowego rumieńca. Dostajesz spokojną, uporządkowaną odpowiedź, która wygląda tak, jakby jej autor właśnie wrócił z biblioteki, gdzie spędził trzy dni na studiowaniu literatury naukowej.

Tymczasem model językowy może stworzyć tekst podobny do prawidłowej bibliografii, nie mając wystarczających podstaw, by potwierdzić prawdziwość każdego elementu. To ważna różnica. Umiejętność napisania wiarygodnie wyglądającego przypisu nie jest tym samym co odnalezienie publikacji, przeczytanie jej i sprawdzenie, czy rzeczywiście potwierdza podane twierdzenie. Podobnie jak umiejętność narysowania mapy skarbów nie oznacza, że pod wskazaną palmą leży skrzynia złota. Czasami leży tam tylko rachunek za wakacje.

A przecież korzystanie z AI jest naprawdę wygodne. Pomaga szybko uporządkować pomysły, zrozumieć skomplikowane zagadnienia, przygotować szkic tekstu czy znaleźć kierunek dalszych poszukiwań. Problem zaczyna się wtedy, gdy wygodę mylimy z dowodem. Odpowiedź brzmi przekonująco, więc przestajemy pytać, skąd pochodzi informacja. Dostajemy link, więc zakładamy, że prowadzi do wiarygodnej treści. Widzimy cudzysłów, więc wierzymy, że przytoczone słowa rzeczywiście padły.

W ten sposób można opublikować cytat, którego nikt nigdy nie wypowiedział, powołać się na nieistniejące badanie albo przekazać współpracownikom błędną informację z dopiskiem „według ekspertów". Eksperci dowiadują się później, że uczestniczyli w badaniu, którego nawet nie widzieli. Być może są równie zaskoczeni jak jednorożec, gdy ktoś wpisuje go do rejestru zwierząt gospodarskich.

Ta książka nie powstała po to, żeby przekonywać cię, że sztuczna inteligencja jest bezużyteczna. Byłoby to równie rozsądne jak wyrzucenie kalkulatora, ponieważ ktoś wpisał niewłaściwe liczby. Chodzi o nauczenie się rozróżniania pomocy w szukaniu informacji od potwierdzenia ich

prawdziwości. Możesz korzystać z AI intensywnie, oszczędzać czas i nadal zachowywać kontrolę nad tym, co ostatecznie uznajesz za fakt.

Najpierw warto zrozumieć pewną niewygodną prawdę: nie wszystkie błędne źródła są całkowicie zmyślone. Gdyby każde fałszywe odwołanie prowadziło do artykułu zatytułowanego „Niezwykle Ważne Badanie, Które Właśnie Wymyśliłem", życie byłoby znacznie prostsze. Częściej spotkasz jednak subtelne pomyłki. Prawdziwa publikacja zostaje przypisana niewłaściwemu autorowi. Rzeczywisty raport otrzymuje nieprawidłowy rok wydania. Cytat jest parafrazą przedstawioną jako dosłowna wypowiedź. Link prowadzi do autentycznej strony, ale zamieszczona tam treść nie potwierdza omawianego wniosku.

Dlatego samo pytanie „Czy to źródło istnieje?" nie wystarcza. Trzeba również wiedzieć, czy wskazano właściwy dokument, czy przytoczono jego rzeczywistą treść, czy autor źródła wypowiada się w danej sprawie oraz czy wyciągnięty wniosek nie wykracza poza dostępne dowody. To kilka różnych czynności, chociaż w pośpiechu łatwo wrzucić je do jednej szuflady oznaczonej „Sprawdzone". Szuflada wygląda porządnie. Jej zawartość już niekoniecznie.

W praktyce będziemy traktować odpowiedź AI jako punkt wyjścia do weryfikacji, a nie jako automatyczny certyfikat prawdy. Nauczysz się rozpoznawać sygnały ostrzegawcze w bibliografiach, wyszukiwać oryginalne dokumenty, kontrolować autorów i daty, sprawdzać identyfikatory publikacji oraz odnajdywać cytaty w ich właściwym kontekście. Zobaczysz również, dlaczego pozornie drobne różnice między „badanie wykazało", „autorzy sugerują" i „w

określonych warunkach zaobserwowano" mogą całkowicie zmienić sens informacji.

Nie oznacza to konieczności zostania zawodowym detektywem bibliograficznym. Nie potrzebujesz gabinetu wyłożonego książkami, lupy ani dramatycznej muzyki podczas otwierania kolejnej zakładki przeglądarki. Potrzebujesz metody pozwalającej ustalić, co należy sprawdzić najpierw, gdzie szukać potwierdzenia i kiedy dostępne dowody są niewystarczające. Inaczej będziesz poświęcać tyle samo czasu na weryfikowanie nazwy mało istotnego artykułu, co na sprawdzanie informacji, od której zależy ważna decyzja.

Nie każda sytuacja wymaga przecież identycznej dokładności. Jeśli przygotowujesz luźną notatkę do własnego użytku, możesz zaakceptować większą niepewność, pod warunkiem że jej nie ukrywasz. Jeśli tworzysz publikację naukową, materiał edukacyjny, dokument zawodowy albo tekst dotyczący zdrowia, prawa czy finansów, konsekwencje błędu bywają poważniejsze. W takich przypadkach sprawdzanie źródeł nie jest przesadną ostrożnością. Jest częścią odpowiedzialności za przekazywaną informację.

Ważna będzie również umiejętność przyznawania, że czegoś nie udało się potwierdzić. Brzmi banalnie, ale współczesne narzędzia przyzwyczaiły nas do natychmiastowych odpowiedzi. Puste miejsce w bibliografii wydaje się niemal osobistą porażką. Tymczasem uczciwe stwierdzenie „Nie znalazłem wiarygodnego potwierdzenia tego cytatu" ma większą wartość niż doskonale sformatowany przypis prowadzący donikąd. Brak dowodu nie zawsze dowodzi fałszu, ale z pewnością nie uprawnia do przedstawiania twierdzenia jako zweryfikowanego.

Z czasem zaczniesz dostrzegać, że najważniejszą umiejętnością nie jest łapanie AI na pomyłkach. Jest nią zadawanie właściwych pytań i odróżnianie odpowiedzi od dowodu. Model może pomóc znaleźć trop, wskazać możliwe publikacje, zaproponować frazy wyszukiwania i uporządkować wyniki. Ostateczne potwierdzenie wymaga jednak sprawdzenia odpowiednich materiałów, najlepiej u pierwotnego wydawcy lub w innym wiarygodnym miejscu udostępniającym dokument.

Jeżeli więc kiedykolwiek skopiowałeś przypis bez otwierania źródła, uwierzyłeś cytatowi tylko dlatego, że brzmiał mądrze, albo uznałeś obecność linku za wystarczający dowód, jesteś we właściwym miejscu. Nie będziemy organizować publicznego procesu twojej przeglądarki. Nauczymy się czegoś znacznie przydatniejszego: korzystania z możliwości AI bez oddawania jej prawa do decydowania, co jest prawdą.

Bo sztuczna inteligencja może błyskawicznie przygotować listę dziesięciu źródeł. Twoim zadaniem nie jest zachwycić się długością listy. Twoim zadaniem jest wiedzieć, ile z tych dziesięciu pozycji rzeczywiście zasługuje na obecność w twoim tekście.

A jeśli okaże się, że tylko jedna istnieje, przynajmniej będziesz wiedzieć, która.

Rozdział 1 - Bibliografia z krainy fantazji, czyli dlaczego AI wymyśla źródła

Poniedziałek, godzina 9:17. W sali konferencyjnej trwa spotkanie, na którym prezentujesz wyniki analizy przygotowanej poprzedniego wieczoru. Wszystko idzie znakomicie. Wykresy są czytelne, wnioski rozsądne, a prezentacja ma nawet slajd ze źródłami, co w świecie korporacyjnych prezentacji bywa traktowane niemal jak przejaw arystokratycznego wychowania. W pewnym momencie ktoś pyta, skąd pochodzi informacja o skuteczności omawianego rozwiązania. Bez wahania wskazujesz publikację naukową, którą dzień wcześniej zaproponowała ci sztuczna inteligencja.

Twój rozmówca wpisuje tytuł artykułu do wyszukiwarki. Nic. Sprawdza nazwisko autora. Znajduje naukowca, ale zajmującego się zupełnie inną dziedziną. Próbuje odnaleźć czasopismo. Ono akurat istnieje, chociaż wskazany numer nie zawiera żadnej podobnej publikacji. W sali robi się cicho. Nawet projektor, dotychczas wydający dźwięki przypominające start niewielkiego samolotu, zdaje się pracować dyskretniej.

Właśnie odkryłeś, że twoja bibliografia posiada więcej wyobraźni niż dział kreatywny.

To wyjątkowo nieprzyjemna sytuacja, ponieważ błąd nie wyglądał jak błąd. Nie było literówek, przypadkowych słów ani podejrzanych zdań. Źródło miało autora, datę, tytuł, nazwę czasopisma i numer tomu. Wszystkie elementy znajdowały się na właściwych miejscach. Zabrakło jedynie drobiazgu: rzeczywistej publikacji, do której można byłoby zajrzeć.

Warto zrozumieć, dlaczego takie sytuacje w ogóle występują. Bez tego łatwo przyjąć jedno z dwóch skrajnych stanowisk. Pierwsze głosi, że sztuczna inteligencja jest genialnym bibliotekarzem i należy jej wierzyć. Drugie przedstawia ją jako cyfrowego oszusta, który świadomie próbuje nas kompromitować. Obie wizje są efektowne, ale nie pomagają dobrze korzystać z narzędzia.

Model językowy nie musi posiadać prawdziwego zamiaru wprowadzenia cię w błąd, żeby wygenerować nieprawdziwą informację. Może po prostu stworzyć odpowiedź, która pasuje do wzorców poznanych podczas treningu, mimo że nie została dostatecznie powiązana z rzeczywistym dokumentem. To zasadnicza różnica między poprawnym wyglądem odpowiedzi a prawdziwością jej zawartości.

Gdy elegancki przypis udaje dowód

Wyobraź sobie zadanie polegające na uzupełnieniu zdania: „Według badań opublikowanych w czasopiśmie...". Człowiek znający język i podstawowe zasady pisania naukowego potrafi zaproponować naturalne zakończenie. Nie oznacza to jednak, że właśnie sprawdził archiwum odpowiedniego wydawnictwa. Podobnie model może tworzyć bibliograficznie przekonujące zestawienia, wykorzystując rozpoznane zależności między nazwiskami, tytułami, dziedzinami i sposobami zapisu publikacji.

W zależności od konstrukcji i dostępnych funkcji narzędzie może również przeszukiwać internet, odczytywać dokumenty lub korzystać z zewnętrznych baz. Takie możliwości poprawiają warunki weryfikacji, lecz same nie gwarantują poprawności końcowej odpowiedzi. Model może źle zinterpretować znaleziony materiał, połączyć informacje z dwóch publikacji albo przypisać rzeczywistemu źródłu wniosek, którego autorzy wcale nie sformułowali.

Problem ma więc kilka poziomów. Czasami źródło nie istnieje. Czasami istnieje, lecz zostało błędnie opisane. Innym razem wszystkie dane bibliograficzne są poprawne, a pomyłka dotyczy znaczenia tekstu. Ta ostatnia sytuacja jest szczególnie zdradliwa, ponieważ zwykłe sprawdzenie, czy publikacja znajduje się w wyszukiwarce, nie ujawnia błędu.

Przyjmijmy fikcyjny przykład. Prosisz AI o badania dotyczące wpływu krótkich przerw na koncentrację podczas pracy biurowej. Otrzymujesz tytuł „The Impact of Microbreaks on Cognitive Performance in Office Environments", nazwiska autorów, rok wydania oraz nazwę czasopisma psychologicznego. Wszystko brzmi niezwykle prawdopodobnie. Temat pasuje do dziedziny, tytuł ma naukową powagę, a nazwiska wyglądają tak, jakby ich właściciele regularnie uczestniczyli w konferencjach.

Nie oznacza to, że wskazany artykuł istnieje. Tytuł w tym przykładzie jest ilustracją, a nie potwierdzoną pozycją bibliograficzną. Właśnie tak należy traktować każdą niezweryfikowaną propozycję źródła: jako informację wymagającą sprawdzenia, nawet jeśli wygląda profesjonalnie.

Co ciekawe, im bardziej szczegółowa odpowiedź, tym silniejsze może wywoływać wrażenie wiarygodności. Numer strony, data publikacji i identyfikator DOI przypominają pieczątki urzędowe. Człowiek patrzy na nie i podświadomie zakłada, że ktoś już wykonał odpowiednią kontrolę. Tymczasem szczegółowość może być jedynie szczegółowością. Bardzo precyzyjny błąd nadal pozostaje błędem, tylko lepiej ubranym.

Nie trzeba zresztą szukać wymyślonego czasopisma, aby wpaść w kłopoty. Wystarczy prawdziwa nazwa renomowanego wydawnictwa połączona z nieistniejącym

artykułem. Autentyczny element nadaje całemu opisowi pozory solidności. To trochę tak, jakby ktoś podał ci fikcyjny adres mieszkania w rzeczywiście istniejącym wieżowcu. Budynek stoi, ulica się zgadza, ale pod wskazanymi drzwiami znajduje się ściana.

Dlaczego człowiek tak łatwo daje się przekonać

Nasze kłopoty nie wynikają wyłącznie z działania technologii. Drugą połowę układanki stanowi sposób, w jaki oceniamy informacje. Kiedy tekst jest płynny, uporządkowany i napisany pewnym tonem, częściej wydaje nam się kompetentny. Szczególnie wtedy, gdy dotyczy tematu, na którym sami nie znamy się wystarczająco dobrze.

Nie pomaga również fakt, że zazwyczaj prosimy AI o wsparcie właśnie wtedy, gdy chcemy zaoszczędzić czas. Potrzebujemy szybkiego materiału, więc otrzymana bibliografia idealnie pasuje do naszych oczekiwań. Weryfikacja oznacza dodatkową pracę. Otwieranie publikacji, sprawdzanie autorów i porównywanie cytatów brzmi znacznie mniej atrakcyjnie niż naciśnięcie przycisku „Kopiuj".

Do tego dochodzi subtelna pokusa potwierdzenia własnego stanowiska. Jeśli narzędzie przedstawia pięć źródeł wspierających naszą tezę, odczuwamy satysfakcję. Gdyby wskazało pięć publikacji podważających tę samą tezę, prawdopodobnie zaczęlibyśmy zadawać dodatkowe pytania. Właśnie dlatego szczególnie uważnie trzeba traktować źródła, które mówią dokładnie to, co chcieliśmy usłyszeć.

Przykładowo przygotowujesz argumentację za wprowadzeniem określonego sposobu organizacji pracy. Pytasz: „Jakie badania dowodzą, że ta metoda zwiększa

produktywność?". Już samo pytanie zakłada, że odpowiednie dowody istnieją. Model może odpowiedzieć zgodnie z tak sformułowanym oczekiwaniem, zamiast zakwestionować jego podstawę.

Rozsądniej zapytać: „Jakie wiarygodne publikacje badają skuteczność tej metody? Wskaż wyniki zarówno korzystne, jak i niekorzystne, oraz wyraźnie zaznacz informacje, których nie możesz potwierdzić". Takie sformułowanie nie zapewni bezbłędności, ale zmienia kierunek poszukiwań. Przestajesz zamawiać dowody na wcześniej przygotowany wniosek.

To ważne także wtedy, gdy korzystasz z narzędzia, które rzeczywiście ma dostęp do wyszukiwarki. Internet zawiera mnóstwo materiałów różnej jakości. Samo odnalezienie strony nie oznacza, że opisane na niej twierdzenie jest prawdziwe. Wyszukiwanie i ocenianie wiarygodności to dwie odrębne czynności.

Cztery postacie nieprawdziwego źródła

W praktyce dobrze rozpoznawać cztery podstawowe sytuacje, ponieważ każda wymaga innej reakcji.

Pierwsza to całkowita fikcja bibliograficzna. Publikacja, autorzy lub czasopismo nie odpowiadają rzeczywistemu dokumentowi. Możesz spędzić pół godziny, szukając artykułu, który istnieje wyłącznie w wygenerowanej odpowiedzi. Jeśli po rozsądnym sprawdzeniu katalogów i baz nie znajdujesz potwierdzenia, nie wpisuj go do bibliografii jako faktu. Nie musisz dowodzić, że dokument nigdy nie istniał w całej historii ludzkości. Wystarczy uczciwie stwierdzić, że nie udało się go zweryfikować.

Druga sytuacja to pomieszanie danych. Artykuł istnieje, lecz podano niewłaściwy rok, tytuł lub autorów. Bywa, że

model składa jedną pozycję bibliograficzną z fragmentów kilku rzeczywistych publikacji. Rozpoznanie tego błędu wymaga porównania pełnego opisu z oryginalnym rekordem bibliograficznym. Nie należy poprawiać poszczególnych elementów na wyczucie, ponieważ można przypadkiem stworzyć kolejną hybrydę.

Trzecia postać dotyczy zniekształconego cytatu. Autor rzeczywiście napisał określone zdanie, ale AI zmienia kilka słów, pomija ważne zastrzeżenie albo łączy dwa fragmenty w jedną wypowiedź. Sens może pozostać podobny, jednak tekst zapisany w cudzysłowie powinien odpowiadać oryginałowi. Jeśli chcesz przedstawić znaczenie wypowiedzi własnymi słowami, zastosuj parafrazę i odpowiednie wskazanie źródła. Nie przebieraj parafrazy za cytat dosłowny.

Czwarta sytuacja jest najbardziej podstępna: źródło istnieje, cytat jest poprawny, lecz nie potwierdza stawianego twierdzenia. Publikacja opisuje związek między dwiema zmiennymi, a AI przedstawia go jako dowód przyczynowości. Badanie dotyczy niewielkiej, specyficznej grupy, a odpowiedź formułuje wniosek dotyczący wszystkich ludzi. Autorzy piszą o możliwości, a wygenerowane streszczenie ogłasza pewność.

Możesz więc mieć prawdziwy artykuł, poprawny link i nadal błędny argument. Bibliografia zostaje uniewinniona, lecz wniosek właśnie uciekł z miejsca zdarzenia.

Pierwszy filtr, zanim zaczniesz szukać

Nie musisz od razu przeprowadzać pełnej analizy każdej publikacji. W codziennej pracy przydaje się wstępny filtr pozwalający zdecydować, które pozycje wymagają dokładniejszego sprawdzenia. Zacznij od przeczytania

odpowiedzi AI z założeniem, że jest propozycją, a nie zakończoną pracą badawczą.

Zwróć uwagę na źródła przedstawione bez pełnych danych, podejrzanie idealnie dopasowane do twojego pytania albo zawierające bardzo stanowcze twierdzenia bez wskazania konkretnych fragmentów dokumentu. Takie cechy nie dowodzą fałszu, lecz stanowią powód do kontroli. Również brak autora czy daty nie zawsze oznacza niewiarygodność, ponieważ nie wszystkie wartościowe materiały mają strukturę publikacji naukowej.

Przy każdej ważnej pozycji spróbuj odpowiedzieć na trzy pytania. Czy potrafię odnaleźć dokument niezależnie od odpowiedzi AI? Czy opis bibliograficzny odpowiada rzeczywistemu dokumentowi? Czy wiem, jaki dokładnie fragment potwierdza moje twierdzenie? Jeśli zatrzymujesz się już na pierwszym pytaniu, nie przechodź automatycznie do trzeciego, udając, że wszystko jest w porządku.

Załóżmy, że otrzymałeś pięć źródeł do prezentacji. Dwa znajdujesz szybko w wiarygodnych katalogach. Jedno posiada błędny rok, ale oryginalny dokument można zidentyfikować. Pozostałych dwóch nie udaje się odnaleźć. Nie oznacza to, że musisz wyrzucić całą pracę. Możesz zachować potwierdzone pozycje, poprawić dane trzeciej i usunąć dwie niezweryfikowane. Ważne, żeby nie traktować wszystkich pięciu jako jednakowo wiarygodnych.

Jeśli źródło jest kluczowe dla ważnej decyzji, samo potwierdzenie jego istnienia będzie dopiero początkiem. Potrzebna stanie się ocena treści i ograniczeń publikacji. Natomiast w przypadku materiału pomocniczego możesz po prostu zrezygnować z niepewnego odwołania, zamiast prowadzić wielogodzinne poszukiwania.

Warto też zachować ślad własnej kontroli. Nie musi to być skomplikowany system. Krótka notatka „potwierdzone u wydawcy", „nie znaleziono pełnego tekstu" albo „źródło nie wspiera wniosku" często wystarcza, by po tygodniu wiedzieć, dlaczego określona publikacja została wykorzystana lub odrzucona.

AI jako pomocnik, nie świadek koronny

Najbardziej użyteczna zmiana polega na rozdzieleniu dwóch ról. Sztuczna inteligencja może znakomicie pomagać w przygotowywaniu pytań, porządkowaniu pojęć i tworzeniu strategii wyszukiwania. Nie powinna jednak automatycznie stawać się ostatecznym świadkiem prawdziwości informacji, którą sama wygenerowała.

Kiedy prosisz: „Czy podane przez ciebie źródła są prawdziwe?", możesz otrzymać kolejną przekonującą odpowiedź, również wymagającą weryfikacji. To nie jest niezależna kontrola. Narzędzie może skorygować błąd, ale może też podtrzymać wcześniejszą nieścisłość. Samo zapewnienie „sprawdziłem" nie zastępuje możliwości wskazania, co faktycznie zostało sprawdzone i gdzie.

Lepiej wykorzystać AI do przygotowania listy elementów wymagających potwierdzenia, a następnie porównać je z oryginalnymi dokumentami. Możesz również poprosić o wskazanie ograniczeń własnej odpowiedzi. Otrzymasz dzięki temu pomoc w organizacji pracy, bez przenoszenia całej odpowiedzialności na system.

Nie trzeba podchodzić do każdego wygenerowanego zdania jak do podejrzanego w sprawie kryminalnej. Wystarczy dostosować staranność do konsekwencji błędu. Informacja użyta jako luźna inspiracja ma inną wagę niż cytat w pracy dyplomowej albo podstawa decyzji

biznesowej. Właśnie w tych ważniejszych przypadkach potrzebujesz czegoś więcej niż eleganckiego przypisu.

Jeżeli więc następnym razem otrzymasz bibliografię wyglądającą jak owoc miesięcznej pracy zespołu badaczy, nie odrzucaj jej automatycznie. Nie przyjmuj jej również bez sprawdzenia. Potraktuj ją jak listę potencjalnych adresów, pod które warto zajrzeć.

Bo najbardziej profesjonalnie sformatowana bibliografia świata nadal nie pomoże, jeśli połowa jej autorów mieszka wyłącznie w wyobraźni modelu.

Rozdział 2 - Cytat w cudzysłowie, prawda na urlopie

Cudzysłów jest jednym z najmniejszych znaków typograficznych, którym ludzie przyznają zaskakująco wielką władzę. Wystarczy umieścić kilka słów między dwiema parami kresek, dopisać nazwisko znanej osoby i nagle zwyczajne zdanie zaczyna wyglądać jak historyczna wypowiedź. Jeśli obok pojawi się data, tytuł książki albo nazwa konferencji, wiarygodność rośnie jeszcze bardziej. Człowiek przestaje pytać, czy ktoś rzeczywiście wypowiedział te słowa, i zaczyna zastanawiać się, jak najlepiej wykorzystać je w swojej prezentacji.

Pewnego dnia przygotowujesz artykuł dotyczący podejmowania decyzji. Prosisz AI o krótki cytat znanego myśliciela, który podkreśla znaczenie krytycznego myślenia. Otrzymujesz zdanie tak trafne, że aż chciałoby się umieścić je na plakacie nad biurkiem. Wypowiedź ma odpowiednio poważny rytm, autor jest rozpoznawalny, a przesłanie idealnie pasuje do tekstu.

Kiedy jednak próbujesz odnaleźć cytat w oryginalnej publikacji, pojawia się problem. Wyszukiwarka pokazuje setki stron powtarzających tę samą wypowiedź, ale żadna nie wskazuje konkretnego miejsca jej pierwszego wystąpienia. Jeden portal podaje książkę, inny przemówienie, trzeci datę, której nie da się połączyć z żadnym potwierdzonym wydarzeniem. Czwarty cytuje pierwszy, piąty czwarty, a szósty zapewnia, że zdanie jest „powszechnie znane”.

Powstała właśnie cyfrowa orkiestra, w której wszyscy muzycy grają z tej samej błędnej kartki.

Problem fałszywych cytatów jest starszy niż sztuczna inteligencja. Ludzie od dawna przypisują znanym postaciom wypowiedzi, które brzmią odpowiednio mądrze, zabawnie albo inspirująco. Internet ułatwił ich rozpowszechnianie, a narzędzia AI mogą dodatkowo przyspieszać tworzenie nowych wersji. Nie każda nieprawidłowo przypisana wypowiedź powstała jednak wskutek działania modelu językowego. Część błędów była obecna w dostępnych materiałach jeszcze przed jego treningiem.

Właśnie dlatego sprawdzanie cytatów wymaga większej precyzji niż zwykłe pytanie o autora. Należy ustalić, jakie słowa rzeczywiście padły, w jakim kontekście, gdzie zostały utrwalone i czy wskazana osoba faktycznie jest ich autorem. Dopiero wtedy cudzysłów zaczyna spełniać swoją funkcję.

Kiedy parafraza zakłada cudzy garnitur

Najczęstszy kłopot nie musi polegać na całkowitym wymyśleniu wypowiedzi. Czasami AI rozpoznaje ogólny sens czyjejś myśli i formułuje go własnymi słowami, a następnie przedstawia jako cytat dosłowny. Powstaje zdanie wierne pewnemu poglądowi, lecz niewierne rzeczywistemu brzmieniu tekstu.

Załóżmy, że autor publikacji pisze: „W analizowanej grupie uczestników regularne przerwy wiązały się z mniejszym deklarowanym zmęczeniem”. AI może przekształcić ten fragment w znacznie bardziej efektowne: „Regularne przerwy zmniejszają zmęczenie pracowników”. Na pierwszy rzut oka oba zdania dotyczą podobnego zagadnienia. Różnica jest jednak ogromna.

Pierwsze opisuje zaobserwowany związek w określonej grupie. Drugie sugeruje ogólną zależność przyczynową. Usunięto ograniczenie dotyczące badanych osób, zniknęła

informacja o deklarowanym charakterze zmęczenia, a ostrożne sformułowanie zastąpiono stanowczym twierdzeniem.

Jeżeli drugie zdanie umieścisz w cudzysłowie i podpiszesz nazwiskiem autora, przypiszesz mu słowa, których w tym przykładzie nie wypowiedział. Co więcej, zmienisz znaczenie jego wypowiedzi. To już nie jest niewinna kosmetyka językowa.

Pamiętaj, że cytat dosłowny i parafraza mają odmienne zadania. Cytat służy wiernemu przytoczeniu konkretnych słów. Parafraza pozwala przedstawić cudzą myśl własnym językiem, przy zachowaniu jej znaczenia i odpowiednim wskazaniu źródła. W obu przypadkach trzeba rozumieć oryginał, ale tylko w pierwszym oczekuje się zgodności przytoczonego brzmienia.

Sztuczna inteligencja może pomóc przygotować parafrazę, pod warunkiem że masz dostęp do tekstu źródłowego i możesz ocenić jej poprawność. Gdy natomiast prosisz o dokładny cytat, potrzebujesz rzeczywistego fragmentu dokumentu. Model nie powinien zastępować brakujących słów własną twórczością, nawet jeśli robi to z literackim wdziękiem.

Przy cytatach tłumaczonych pojawia się dodatkowa trudność. Polski przekład angielskiej wypowiedzi nie jest dosłownym zapisem oryginalnych słów w tym samym języku. Można go przytoczyć jako tłumaczenie, ale należy odpowiednio zaznaczyć jego charakter. Jeśli korzystasz z opublikowanego przekładu, warto wskazać jego źródło. Jeżeli tłumaczysz samodzielnie, nie przedstawiaj własnej wersji jako jedynego autoryzowanego brzmienia.

Różnica ma znaczenie zwłaszcza w tekstach naukowych i zawodowych, w których jedno słowo może zmieniać zakres twierdzenia. Angielskie „may" nie jest tym samym co „must", a „associated with" nie oznacza automatycznie „caused by". Wygładzenie języka potrafi niepostrzeżenie zamienić ostrożny wniosek w kategoryczne zapewnienie.

Siedemset stron powtarza to samo. Czyli kto sprawdził?

Jedną z najbardziej zwodniczych cech internetu jest możliwość wielokrotnego powielenia informacji bez zwiększenia liczby niezależnych dowodów. Cytat może pojawiać się na setkach stron, w tysiącach postów i dziesiątkach prezentacji, choć wszystkie te materiały pochodzą ostatecznie z jednego niezweryfikowanego wpisu.

Wyobraź sobie, że ktoś publikuje efektowne zdanie przypisane znanemu naukowcowi. Kolejny autor kopiuje je do artykułu, następny umieszcza na grafice, a jeszcze inny dodaje do internetowej kolekcji inspirujących cytatów. Po kilku latach wyszukiwarka zwraca mnóstwo wyników. Samo powtórzenie informacji sprawia, że zaczyna ona wyglądać znajomo, a znajomość bywa mylona z prawdziwością.

Nie wystarczy więc policzyć stron zawierających określone słowa. Trzeba ustalić, czy którakolwiek z nich wskazuje wiarygodne źródło pierwotne.

W przypadku wypowiedzi autora książki może to być konkretne wydanie wraz z numerem strony lub innym jednoznacznym miejscem w tekście. W przypadku publicznego przemówienia przydatny będzie oficjalny zapis, nagranie lub wiarygodna transkrypcja. Jeśli chodzi o artykuł naukowy, najlepiej dotrzeć do jego właściwej treści, a nie zatrzymywać się na omówieniu przygotowanym przez inną osobę.