PROBLEM? WYLUZUJ. OGARNIESZ TO!

Napisałem To Sam, Ale Program Twierdzi, Że Jestem Robotem

Jak dokumentować własny proces tworzenia, spokojnie odpowiadać na zarzut użycia AI i przedstawiać dowody autorstwa bez uznawania wyniku detektora za ostateczny wyrok



Max Paradox

Napisałem To Sam, Ale Program Twierdzi, Że Jestem Robotem

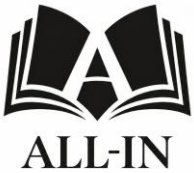
Jak dokumentować własny proces tworzenia, spokojnie odpowiadać na zarzut użycia AI i przedstawiać dowody autorstwa bez uznawania wyniku detektora za ostateczny wyrok

Max Paradox

PROBLEM? WYLUZUJ. OGARNIESZ TO!

ISBN: 9788384839492

ALL-IN Rafał Płonkowski



SPIS TREŚCI

INTRO 5

Rozdział 1 - Kiedy własny tekst staje się podejrzany 10

Rozdział 2 - Co detektor naprawdę widzi, a czego nigdy się nie dowie 18

Rozdział 3 - Dlaczego nagle piszesz podejrzanie dobrze? 26

Rozdział 4 - Jak nie zniszczyć własnej wiarygodności podczas jej ratowania 34

Rozdział 5 - Twoja historia pisania nie powinna znikać razem z przyciskiem „Zapisz” 42

Rozdział 6 - Kiedy sztuczna inteligencja pomaga, a kiedy zaczyna pisać za ciebie 50

Rozdział 7 - Cudze słowa w twoim tekście, czyli skąd naprawdę pochodzi treść 58

Rozdział 8 - Nie musisz wygrać kłótni, żeby skutecznie wyjaśnić autorstwo 66

Rozdział 9 - Zbuduj teczkę dowodową, która potrafi mówić bez ciebie 74

Rozdział 10 - Napisz odpowiedź, której nie trzeba tłumaczyć drugi raz 82

Rozdział 11 - Weź jeden akapit pod lupę i pokaż, jak naprawdę powstał 91

Rozdział 12 - Odwołanie bez wojny, czyli co zrobić, kiedy decyzja już zapadła 99

Rozdział 13 - Ustal zasady autorstwa, zanim ktoś zacznie szukać robota 108

Rozdział 14 - Jak sprawdzać autorstwo, nie zamieniając oceny w polowanie na czarownice 117

Rozdział 15 - Kiedy oskarżenie wychodzi poza ekran i zaczyna żyć własnym życiem 127
Rozdział 16 - Kiedy dowody zaczynają sobie przeczyć 136
Rozdział 17 - Mam dowody, ale nie mogę ich pokazać. I co teraz? 145
Rozdział 18 - System, który pamięta za ciebie, kiedy ty zajmujesz się pisaniem 153
Rozdział 19 - Kiedy wraca stary lęk, czyli jak nie zostać zakładnikiem detektora 162
Rozdział 20 - Nie potrzebujesz certyfikatu człowieczeństwa 170

INTRO

Wyobraź sobie, że właśnie skończyłeś pisać tekst. Poświęciłeś mu kilka wieczorów, trzy kubki kawy, siedem zmian koncepcji i jedną poważną rozmowę z samym sobą na temat sensu życia. Poprawiłeś przecinki, wyrzuciłeś połowę przymiotników i uznałeś, że ostatnie zdanie wreszcie brzmi jak napisane przez człowieka, który wie, co robi. Wysyłasz gotowy materiał. Następnego dnia dostajesz wiadomość: „System wykrył 94% prawdopodobieństwa użycia sztucznej inteligencji". Patrzysz na ekran. Patrzysz na swoje ręce. Sprawdzasz puls. Wszystko wskazuje na to, że jesteś człowiekiem, ale najwyraźniej komputer ma inne zdanie. Być może powinieneś zgłosić się do producenta z reklamacją własnego człowieczeństwa.

Najbardziej absurdalne jest to, że jeszcze niedawno poprawne pisanie uchodziło za zaletę. Staranna konstrukcja zdań, logiczny układ argumentów i konsekwentny styl były dowodami warsztatu. Teraz te same cechy mogą wzbudzać podejrzenia, ponieważ przypominają wzorce spotykane również w tekstach generowanych automatycznie. Oczywiście nie każdy detektor działa identycznie, a jego wynik zależy od zastosowanego rozwiązania i analizowanego materiału. Problem pojawia się wtedy, gdy odbiorca traktuje liczbowy rezultat jako bezpośredni dowód pochodzenia tekstu. Człowiek spędził tydzień nad artykułem, algorytm przeanalizował go w kilka sekund, a następnie ktoś uznał, że to algorytm lepiej pamięta, kto siedział przy klawiaturze. Trudno wymyślić bardziej ekonomiczny sposób na doprowadzenie autora do szału.

Zarzut użycia AI potrafi zaboleć mocniej, niż sugerowałaby sama treść wiadomości. Nie chodzi przecież wyłącznie o technologię. Ktoś kwestionuje twoją uczciwość, samodzielność albo kompetencje. Student może obawiać się konsekwencji

akademickich, autor utraty wiarygodności, pracownik problemów zawodowych, a freelancer odmowy zapłaty za wykonaną pracę. Nawet jeśli sprawa zakończy się wyjaśnieniem, samo podejrzenie potrafi zmienić zwykłą rozmowę o jakości tekstu w przesłuchanie przypominające komisję śledczą. Brakuje tylko lampy skierowanej w twarz i pytania, gdzie przebywałeś w chwili napisania trzeciego akapitu. W takiej sytuacji zdenerwowanie jest zrozumiałe, ale nie zawsze pomaga w obronie własnej pracy.

Pierwszy odruch bywa bardzo ludzki: natychmiast udowodnić wszystkim, że się mylą. Otwierasz pięć kolejnych detektorów, wklejasz identyczny tekst i otrzymujesz pięć różnych wyników. Jeden program twierdzi, że jesteś człowiekiem, drugi widzi maszynę, trzeci nie jest pewien, a czwarty prawdopodobnie zastanawia się nad zmianą zawodu. Następnie zaczynasz przepisywać fragmenty, dodawać niezgrabne sformułowania i celowo psuć zdania, żeby wynik wyglądał bardziej „ludzko". W pewnym momencie tekst przypomina wypowiedź człowieka, który właśnie nauczył się polskiego, ale robi to pod wodą. Co gorsza, nadal nie masz dowodu autorstwa. Masz tylko tekst zmieniony pod kaprysy narzędzia, którego działanie nie rozstrzyga pierwotnego pytania.

Właśnie dlatego potrzebujemy rozróżnienia, które stanie się fundamentem tej książki: ocena statystyczna tekstu nie jest tym samym co udokumentowana historia jego powstawania. Detektor analizuje pewne właściwości gotowego materiału i na ich podstawie formułuje ocenę. Nie obserwował twojej pracy, nie widział notatek, nie zna kolejnych wersji i nie uczestniczył w podejmowaniu decyzji redakcyjnych. Nie oznacza to, że wszystkie takie narzędzia są bezwartościowe albo że każdą wątpliwość należy zignorować. Oznacza natomiast, że wynik wymaga właściwej interpretacji, znajomości ograniczeń metody i uwzględnienia innych informacji. Liczba wyglądająca naukowo nie

staje się automatycznie rozstrzygnięciem. Nawet jeśli została pokazana wielką czcionką i otoczona czerwonym prostokątem.

Jest jednak druga strona tego problemu, o której trzeba powiedzieć równie wyraźnie. Samo zapewnienie „napisałem to sam" także nie stanowi niezależnego dowodu. Możesz mówić prawdę, a mimo to nie mieć niczego, co pozwoli drugiej osobie sprawdzić przebieg pracy. Jeżeli wszystkie notatki zostały usunięte, wcześniejsze wersje nadpisane, a jedyny zachowany plik nosi nazwę „final_ostateczny_NAPRAWDE_FINAL.docx", odtworzenie procesu będzie trudniejsze. Nie dlatego, że jesteś podejrzany, lecz dlatego, że końcowy dokument pokazuje rezultat, a nie drogę do niego. Dokumentowanie autorstwa wymaga więc czegoś więcej niż wiary we własną pamięć. Dobra wiadomość jest taka, że nie potrzebujesz prywatnego laboratorium kryminalistycznego ani informatyka, który będzie stał za tobą podczas każdego naciśnięcia klawisza.

Najważniejsze ślady twórczej pracy często już istnieją. Mogą to być szkice, notatki, historia wersji dokumentu, wcześniejsza korespondencja, plan tekstu, materiały źródłowe albo zapis zmian wprowadzanych podczas redakcji. Każdy z tych elementów ma swoje ograniczenia. Data modyfikacji pliku nie dowodzi samodzielnego napisania całej zawartości, a pojedynczy zrzut ekranu można wyrwać z kontekstu. Jednak kilka spójnych, autentycznych materiałów potrafi znacznie lepiej wyjaśnić przebieg pracy niż desperackie powtarzanie, że „przecież naprawdę to pisałem". Nauczysz się odróżniać ślady wartościowe od pozornie imponujących, zachowywać je bez przesadnej biurokracji i przedstawiać w sposób zrozumiały dla osoby, która nie zna twojego warsztatu.

Równie istotna będzie umiejętność odpowiadania na zarzuty bez zamieniania sprawy w osobistą wojnę z technologią. Gdy ktoś przesyła ci wynik detektora, nie musisz od razu przygotowywać czterdziestostronicowego manifestu przeciwko algorytmom. Warto

najpierw ustalić, czego dokładnie dotyczy podejrzenie, na jakiej podstawie zostało sformułowane i jaka procedura obowiązuje w danej sytuacji. Inaczej rozmawia się z redaktorem zgłaszającym wątpliwości, inaczej z klientem kwestionującym sposób wykonania zlecenia, a jeszcze inaczej z instytucją stosującą formalne zasady dotyczące korzystania z AI. Spokojna odpowiedź nie oznacza przyjęcia zarzutu. Oznacza zachowanie kontroli nad faktami, zakresem wyjaśnień i własnymi słowami, zanim emocje napiszą za ciebie wiadomość, której później będziesz żałować.

Nie będziemy też udawać, że świat dzieli się na niewinnych autorów i złowrogie maszyny. Rzeczywiste procesy twórcze są bardziej złożone. Ktoś pisze całkowicie samodzielnie, ktoś korzysta z automatycznej korekty, ktoś używa narzędzia do uporządkowania notatek, a ktoś inny generuje obszerne fragmenty i później je redaguje. Te sytuacje nie są automatycznie równoważne, a ich dopuszczalność zależy od zasad konkretnego zadania, umowy, instytucji lub publikacji. Istotne będzie uczciwe określenie, co rzeczywiście zrobiłeś, jakie narzędzia wykorzystałeś i jaki był twój wkład. Nie chodzi o ukrywanie dopuszczalnej pomocy technologicznej ani przedstawianie pracy wygenerowanej jako własnoręcznie napisanej. Chodzi o rzetelne opisanie faktów i niedopuszczenie do sytuacji, w której domysł zostaje pomylony z ustaleniem.

Ta książka nie nauczy cię oszukiwania detektorów ani produkowania fikcyjnych dowodów. Podpowie natomiast, jak tworzyć wiarygodną dokumentację na bieżąco, a także co można uczciwie zrobić, gdy wcześniej nie pomyślałeś o archiwizowaniu swojej pracy. Dowiesz się, dlaczego rekonstruowanie procesu z autentycznych materiałów różni się od dopisywania sobie nieistniejącej historii. Zobaczysz również, kiedy wystarczy rzeczowe wyjaśnienie, kiedy warto poprosić o ponowną ocenę i kiedy sprawa jest na tyle poważna, że potrzebna będzie pomoc odpowiedniej osoby lub specjalisty. Nie obiecuję, że każdy

oskarżyciel przyzna ci rację. Nie wszyscy ludzie zmieniają zdanie pod wpływem faktów, a algorytmy nie zostały wyposażone w przycisk „przepraszam, poniosło mnie".

Najbardziej chciałbym jednak, żebyś przestał traktować pojedynczy wynik jako wyrok dotyczący twojej wartości i uczciwości. Jeżeli stworzyłeś coś samodzielnie, masz prawo spokojnie wyjaśnić, jak to powstało. Jeżeli korzystałeś z narzędzi, możesz opisać ich rzeczywistą rolę bez dopasowywania opowieści do oczekiwań innych. A jeżeli ktoś kwestionuje twoją pracę, możesz oczekiwać sensownego uzasadnienia i możliwości odniesienia się do przedstawionych wątpliwości, w granicach obowiązujących zasad. Najlepsza obrona autorstwa zaczyna się znacznie wcześniej niż w chwili oskarżenia. Zaczyna się wtedy, gdy traktujesz własny proces twórczy jako coś, co warto rozumieć i zachowywać, zamiast pozostawiać po nim wyłącznie gotowy plik.

Być może pewnego dnia technologia będzie potrafiła znacznie dokładniej ustalać pochodzenie treści. Na razie jednak warto pamiętać o prostej rzeczy: komputer może oceniać tekst, ale nie powinien zastępować rzetelnego ustalania faktów. Twoim zadaniem nie jest przekonać każdą aplikację na świecie, że zasługujesz na certyfikat człowieka. Twoim zadaniem jest wiedzieć, co stworzyłeś, potrafić to uczciwie udokumentować i umieć o tym rozmawiać bez paniki. A jeśli jakiś program nadal będzie twierdził, że jesteś robotem, możesz potraktować to jako jego opinię. Tylko nie zaczynaj specjalnie robić literówek, żeby udowodnić, że masz duszę. Twoja klawiatura naprawdę nie musi cierpieć z tego powodu.

Rozdział 1 - Kiedy własny tekst staje się podejrzany

Wiadomość przyszła w czwartek o 16:42. Marta, studentka kończąca pracę semestralną, przeczytała ją raz, potem drugi, a następnie trzeci, licząc na to, że podczas poprzednich dwóch podejść źle zrozumiała język polski. Prowadzący napisał, że przesłany esej wykazuje cechy tekstu wygenerowanego przez sztuczną inteligencję i wymaga dodatkowych wyjaśnień. Nie wskazał konkretnych fragmentów, nie opisał sposobu oceny, nie podał nawet, czy chodzi o przypuszczenie, czy formalny zarzut. Marta wiedziała natomiast doskonale, ile czasu poświęciła na przygotowanie materiału. Miała za sobą cztery wieczory pracy, notatki rozrzucone między zeszytem a komputerem i znacznie większą wiedzę o omawianym zagadnieniu, niż kiedykolwiek zamierzała zdobyć. Teraz siedziała przed ekranem, zastanawiając się, czy powinna odpowiedzieć grzecznie, stanowczo, czy może wysłać zdjęcie siebie piszącej pracę wraz z aktualną gazetą, jak przy negocjacjach z porywaczami.

To oczywiście przykładowa sytuacja, ale zawiera mechanizm, który może dotknąć każdego autora. Wystarczy, że osoba oceniająca tekst nabierze podejrzeń dotyczących sposobu jego powstania. Czasami podstawą będzie wynik automatycznego narzędzia, czasami nagła zmiana stylu, a niekiedy wyłącznie przekonanie, że dany człowiek „nie mógł napisać czegoś takiego”. Ostatni przypadek jest szczególnie interesujący, ponieważ komplement zostaje w nim przebrany za oskarżenie. Napisałeś zbyt dobrze jak na siebie, więc najwyraźniej ktoś ci pomagał. Gdyby tę logikę zastosować konsekwentnie, każda poprawa umiejętności wymagałaby dochodzenia. Na szczęście ocena autorstwa nie musi sprowadzać się do zgadywania, a dobrze przeprowadzona

rozmowa pozwala oddzielić rzeczywiste wątpliwości od luźnych skojarzeń.

Zanim jednak zaczniesz się bronić, musisz wiedzieć, przed czym właściwie masz się bronić. Stwierdzenie „tekst wygląda na wygenerowany" nie oznacza jeszcze tego samego co „złamałeś obowiązujące zasady". Pierwsze może być obserwacją, choćby nieprecyzyjną. Drugie jest zarzutem określonego zachowania. Trzecia możliwość pojawia się wtedy, gdy ktoś informuje cię wyłącznie o wyniku narzędzia i oczekuje wyjaśnienia. Te sytuacje wymagają innych odpowiedzi, ale zdenerwowany autor często wrzuca je do jednego worka oznaczonego napisem „NIESPRAWIEDLIWOŚĆ". Worek szybko robi się ciężki, szczególnie gdy zaczynasz dopisywać do niego przewidywane konsekwencje, których nikt jeszcze nie zapowiedział.

Wróćmy do Marty. Jej pierwszym pomysłem było przygotowanie długiej wiadomości zaczynającej się od zdania: „Jestem oburzona, że w ogóle można mnie o coś takiego podejrzewać". W następnych akapitach planowała przedstawić historię swojej edukacji, stosunek do uczciwości akademickiej oraz opinię o współczesnej technologii. Problem polegał na tym, że żaden z tych elementów nie odpowiadał na podstawowe pytanie: czego prowadzący potrzebował, aby ocenić sposób powstania pracy? Emocjonalna odpowiedź mogłaby być zrozumiała, ale jednocześnie zamieniłaby prośbę o wyjaśnienia w konflikt o wzajemny szacunek. Tymczasem Marta miała pod ręką coś bardziej użytecznego niż rozbudowaną deklarację moralną: rzeczywiste materiały pozwalające pokazać, jak rozwijała tekst.

Podejrzenie, zarzut i dowód to trzy różne rzeczy

Najważniejszą umiejętnością na początku takiej sytuacji jest rozpoznanie statusu informacji, którą otrzymałeś. Podejrzenie oznacza, że ktoś dostrzegł okoliczność wymagającą sprawdzenia. Zarzut oznacza przypisanie ci określonego działania, zwykle sprzecznego z zasadami lub oczekiwaniami. Dowód jest natomiast

informacją, której wiarygodność i znaczenie można ocenić w odniesieniu do konkretnego twierdzenia. Te pojęcia nie stają się synonimami tylko dlatego, że zostały umieszczone w jednej wiadomości. Jeżeli odbiorca pisze, że system oznaczył twoją pracę jako potencjalnie wygenerowaną, wciąż potrzebujesz wiedzieć, jakie wnioski z tego wyciąga. Czy prosi o wyjaśnienie? Czy wstrzymuje ocenę? Czy uruchomiono oficjalne postępowanie? Dopiero odpowiedź pozwala określić następny rozsądny krok.

Warto przy tym pamiętać, że człowiek zgłaszający zastrzeżenia nie zawsze działa ze złej woli. Nauczyciel może być zobowiązany do sprawdzania samodzielności prac, redaktor odpowiada za standardy publikacji, a klient może mieć w umowie warunek dotyczący sposobu przygotowania materiałów. Nieporozumienie pojawia się wtedy, gdy narzędzie pomocnicze zaczyna zastępować ocenę faktów albo gdy zasady są tak nieprecyzyjne, że nawet osoba je egzekwująca nie potrafi powiedzieć, czego dokładnie zabraniają. Uznanie prawa drugiej strony do zadawania pytań nie wymaga zgody na dowolny sposób wyciągania wniosków. Możesz jednocześnie szanować procedurę i oczekiwać, że zostanie przeprowadzona rzetelnie.

Dlatego pierwsza praktyczna czynność jest mniej widowiskowa, niż podpowiada adrenalina. Przeczytaj otrzymaną wiadomość i oddziel to, co rzeczywiście w niej napisano, od własnych przewidywań. Ustal, kto zgłosił zastrzeżenia, czego dotyczą, czy wskazano konkretną podstawę oraz czy określono termin odpowiedzi. Jeżeli czegoś brakuje, nie musisz zgadywać. Możesz poprosić o doprecyzowanie. Zdanie „Proszę o wskazanie fragmentów budzących wątpliwości oraz zasad, na podstawie których oceniany jest sposób przygotowania pracy” jest znacznie bardziej użyteczne niż „Jak państwo śmieją?”. To drugie wprawdzie świetnie prezentuje się w historycznych dramatach, ale rzadko usprawnia wymianę dokumentów.

Istotne jest również ustalenie stawki. Nie każda wątpliwość dotycząca tekstu ma taką samą wagę. Nieformalna uwaga redaktora może prowadzić do zwykłej rozmowy, podczas gdy oficjalny zarzut naruszenia regulaminu uczelni może wymagać przestrzegania określonych terminów i zasad postępowania. Jeżeli sprawa dotyczy wynagrodzenia, reputacji zawodowej albo odpowiedzialności wynikającej z umowy, nie należy działać tak, jakby chodziło o internetową sprzeczkę. W poważniejszych sytuacjach warto skorzystać z właściwej procedury odwoławczej, pomocy przedstawiciela instytucji albo porady prawnej. Ta książka pomoże ci uporządkować informacje i argumenty, ale nie zastępuje analizy konkretnego regulaminu czy przepisów.

Najgorsze decyzje podejmuje się w pierwszym przypływie paniki

Jednym z najbardziej kosztownych odruchów jest natychmiastowe przerabianie kwestionowanego dokumentu. Autor chce pokazać tekst w lepszym świetle, więc zaczyna wymieniać zdania, zmieniać kolejność akapitów i poprawiać miejsca, które jego zdaniem mogły wzbudzić podejrzenia. W rezultacie po godzinie posiada nowy dokument, który różni się od wersji będącej przedmiotem oceny. Jeśli jednocześnie nadpisał pierwotny plik, sam utrudnił sobie wyjaśnienie sprawy. Nie dlatego, że poprawki są czymś podejrzanym. Po prostu przedstawienie historii konkretnego materiału staje się trudniejsze, gdy materiał ten został zastąpiony innym.

Zdecydowanie rozsądniej jest zachować oryginalną wersję, dokładnie taką, jaka została wysłana lub opublikowana. Jeśli istnieją wcześniejsze szkice, również ich nie zmieniaj. Nie poprawiaj dat, nie odtwarzaj rzekomo zagubionych etapów i nie próbuj tworzyć idealnie wyglądającej historii pracy. Prawdziwy proces twórczy bywa chaotyczny. Jeden akapit powstaje od początku do końca, następny przechodzi sześć poprawek, a trzeci pojawia się nagle po długim spacerze. Taki przebieg może wyglądać

mniej elegancko niż uporządkowana prezentacja, ale nie musi być przez to mniej wiarygodny. Dokumentacja nie jest konkursem na najładniejszą chronologię. Jej zadaniem jest odzwierciedlanie rzeczywistych zdarzeń.

Podobnie ostrożnie należy podchodzić do publicznego nagłaśniania sprawy. Gdy czujesz się niesprawiedliwie potraktowany, publikacja zrzutu ekranu i emocjonalnego komentarza może wydawać się doskonałym sposobem na odzyskanie kontroli. Internet prawdopodobnie dostarczy ci poparcia, kilku ekspertów od wszystkiego i jednego człowieka, który uzna, że problem wynika z niewłaściwego ustawienia monitora. Nie zagwarantuje jednak rozwiązania sporu. Publiczne oskarżenia mogą zaostrzyć konflikt, ujawnić poufne informacje albo skomplikować formalne wyjaśnienia. Zanim zdecydujesz się na taki krok, sprawdź, czy istnieje bezpośrednia droga rozmowy i czy nie zaszkodzisz własnej sprawie.

Jak odpowiedzieć, żeby nie brzmieć jak oskarżony we własnym procesie

Dobra pierwsza odpowiedź może być krótka. Powinna potwierdzać otrzymanie informacji, przedstawiać twoje rzeczywiste stanowisko i wskazywać gotowość do merytorycznego wyjaśnienia sprawy. Jeśli tekst napisałeś samodzielnie, możesz powiedzieć to wprost. Nie musisz dodawać siedmiu przysiąg, powoływać się na honor przodków ani przekonywać odbiorcy, że od dzieciństwa gardzisz automatyzacją. Wystarczy na przykład: „Potwierdzam, że samodzielnie przygotowałem wskazany tekst. Mogę przedstawić dostępne notatki i wcześniejsze wersje dokumentu. Proszę o doprecyzowanie podstaw zgłoszonych wątpliwości oraz sposobu ich weryfikacji”. Taka odpowiedź nie rozwiązuje automatycznie sprawy, ale kieruje rozmowę ku informacjom, które można sprawdzić.

Co jednak zrobić, jeśli rzeczywiście korzystałeś z jakiegoś narzędzia AI? Przede wszystkim opisać prawdę bez

wyolbrzymiania i bez ukrywania istotnych elementów. Automatyczne sprawdzenie pisowni, wygenerowanie propozycji struktury czy przygotowanie fragmentu tekstu to różne rodzaje pomocy. Ich znaczenie zależy od faktycznego zakresu wykorzystania i obowiązujących zasad. Nie warto mówić „nie używałem żadnej sztucznej inteligencji", jeśli korzystałeś z niej podczas przygotowania materiału, nawet gdy uważasz taki udział za niewielki. Równie nierozsądne byłoby przyjmowanie zarzutu niedozwolonego generowania całej pracy tylko dlatego, że użyłeś funkcji pomocniczej, która była dozwolona. Precyzja jest tutaj korzystniejsza niż dramatyczne deklaracje.

Czasami największą trudnością jest brak dokumentów. Możliwe, że pisałeś w jednym pliku, nie zachowałeś wersji roboczych, a notatki powstawały na papierze, który później wyrzuciłeś. Nie oznacza to automatycznie, że nie masz żadnej możliwości wyjaśnienia autorstwa. Możesz przedstawić to, co rzeczywiście pamiętasz, wskazać wykorzystane źródła, omówić argumentację i wyjaśnić decyzje dotyczące treści. Możesz również sprawdzić, czy używany program zachował historię dokumentu, choć nie należy zakładać, że taka funkcja była aktywna. Brak śladów trzeba jednak uczciwie nazwać. Nie zastępuj go sztucznie przygotowanymi materiałami. Wiarygodność opiera się także na gotowości do powiedzenia: „Tego elementu nie mogę udokumentować".

Ważne, aby nie wpaść w drugą skrajność i nie próbować przekazać odbiorcy wszystkiego, co kiedykolwiek powstało na twoim komputerze. Setki plików, prywatne wiadomości i pełny dostęp do konta nie stają się automatycznie uzasadnionym wymaganiem tylko dlatego, że ktoś ma wątpliwości dotyczące tekstu. Materiały powinny odpowiadać konkretnemu pytaniu, a ich zakres musi uwzględniać prywatność, poufność i prawa innych osób. Jeżeli ktoś żąda danych wykraczających poza potrzebę weryfikacji, możesz zapytać o podstawę, cel i bezpieczny sposób ich

udostępnienia. Współpraca przy wyjaśnianiu autorstwa nie oznacza rezygnacji z ochrony własnych informacji.

Zdarza się również, że odbiorca nie chce rozmawiać o procesie tworzenia, ponieważ uważa wynik detektora za wystarczający. Wtedy dalsze zapewnienia o uczciwości mogą prowadzić donikąd. Lepiej poprosić o wskazanie przyjętych zasad oceny, zakresu możliwych wyjaśnień i sposobu rozpatrywania materiałów dodatkowych. Jeżeli istnieje formalna procedura, korzystaj z niej i zachowuj korespondencję. Nie musisz przekonywać rozmówcy podczas jednej wymiany wiadomości. W sprawach wymagających rozstrzygnięcia liczy się również możliwość prześledzenia, jakie informacje przedstawiono i jak na nie odpowiedziano. Spokojna konsekwencja bywa skuteczniejsza od coraz głośniejszych protestów, chociaż zdecydowanie gorzej nadaje się na widowiskowy finał filmu.

Twoja praca zasługuje na ocenę, nie na zgadywanie

Najtrudniejszą częścią takiej sytuacji bywa zaakceptowanie, że nawet rzeczowa odpowiedź nie zapewni natychmiastowego uznania twoich argumentów. Druga osoba może potrzebować czasu, może mieć ograniczone możliwości decyzyjne albo może po prostu nie zgodzić się z twoją interpretacją. Nie należy mylić braku szybkiej odpowiedzi z potwierdzeniem winy. Jeżeli dostarczyłeś informacje, zachowałeś terminy i jasno określiłeś swoje stanowisko, wykonałeś konkretną część pracy potrzebnej do wyjaśnienia sprawy. Dalszy przebieg zależy również od procedury i działań innych osób. Nie wszystko da się przyspieszyć przez wysyłanie kolejnej wiadomości co dwadzieścia minut.

Możesz natomiast wyciągnąć użyteczny wniosek na przyszłość. Zamiast traktować dokumentowanie procesu twórczego jako awaryjną czynność wykonywaną po oskarżeniu, warto uznać je za naturalny element pracy. Nie chodzi o gromadzenie nieograniczonej liczby plików ani tworzenie prywatnego archiwum przypominającego magazyn tajnych służb. Chodzi o pozostawianie

autentycznych śladów najważniejszych etapów: pomysłu, szkicu, rozwijania treści i redakcji. Takie materiały pomagają nie tylko w sporach o autorstwo. Ułatwiają również powrót do wcześniejszych pomysłów, wyjaśnianie zmian i ocenę własnego postępu. Ich wartość istnieje niezależnie od tego, czy kiedykolwiek ktoś zakwestionuje twoją pracę.

Marta ostatecznie nie wysłała swojego manifestu. Zachowała otrzymaną wiadomość, sprawdziła dostępne wcześniejsze wersje eseju i poprosiła prowadzącego o wskazanie podstaw wątpliwości. To nie jest opowieść o cudownym zwycięstwie nad systemem. Nie znamy decyzji prowadzącego, bo cała sytuacja jest przykładem ilustrującym właściwy sposób reagowania. Najważniejsze wydarzyło się wcześniej: Marta przestała próbować udowodnić własną uczciwość samą siłą oburzenia i zaczęła porządkować rzeczywiste informacje. Zarzut autorstwa może być poważną sprawą, ale nie wymaga od ciebie natychmiastowego zamieniania się w oskarżyciela, obrońcę i sędziego jednocześnie. Wystarczy rozpocząć od faktów. To znacznie mniej efektowne niż dramatyczne przemówienie, za to zdecydowanie bardziej przydatne.

Rozdział 2 - Co detektor naprawdę widzi, a czego nigdy się nie dowie

Na ekranie pojawia się wynik: 87%. Obok znajduje się pomarańczowy pasek, ostrzegawczy symbol i komunikat sugerujący, że analizowany tekst mógł powstać przy udziale sztucznej inteligencji. Wszystko wygląda profesjonalnie. Interfejs jest uporządkowany, liczba precyzyjna, a grafika sprawia wrażenie przygotowanej przez ludzi, którzy wiedzą znacznie więcej od przeciętnego użytkownika. Wystarczy jednak zadać jedno niewygodne pytanie: 87% czego? Prawdopodobieństwa, że cały tekst napisała maszyna? Udziału fragmentów przypominających treści generowane? Pewności wewnętrznego klasyfikatora? A może wartości wskaźnika, którego dokładnego znaczenia producent nie wyjaśnia? Jeżeli odpowiedź nie jest jasna, sama liczba nie rozwiązuje problemu. Nawet jeśli wygląda tak przekonująco, że człowiek odruchowo chce zapytać, gdzie może zapłacić mandat.

Żeby właściwie rozumieć działanie detektorów, trzeba zacząć od ich podstawowego ograniczenia. Typowy detektor analizuje gotowy materiał, szukając cech pozwalających odróżniać określone grupy tekstów. Nie widzi autora przy biurku, nie zna przebiegu jego myślenia i nie obserwuje procesu powstawania kolejnych zdań. Może wykorzystywać model statystyczny, klasyfikator uczony na przykładach albo zestaw różnych technik analitycznych. Szczegółowe rozwiązania zależą od producenta i nie zawsze są publicznie opisane. Wspólna pozostaje jednak zasadnicza różnica między analizą właściwości tekstu a bezpośrednim ustaleniem jego pochodzenia. Pierwsza czynność może dostarczyć wskazówek. Druga wymaga informacji o tym, co naprawdę wydarzyło się podczas tworzenia materiału.

Statystyka nie prowadzi dziennika twojego życia