Martyna Tobiasz

PRACOWNICY GLOBALNEGO POŁUDNIA JAKO OFIARY ROZWOJU SZTUCZNEJ INTELIGENCJI

Gdy świat zachwyca się kolejnymi przełomami technologicznymi, a mocarstwa rywalizują w wyścigu o jej rozwój, w państwach Globalnego Południa bezimienni pracownicy filtrują miliony danych, by uczyć używane na co dzień przez ludzi modele językowe odróżniania prawdy od fałszu i budowania zdań, które zadowolą jak największą ilość osób. Przypisując sobie całe zasługi rozwoju tej technologii i generując miliardowe zyski, korporacje technologiczne wykorzeniane na Globalnej Północy wykorzystują nierówności strukturalne i dysproporcje ekonomiczne, przenosząc procesy o wysokim obciążeniu psychicznym i niskim statusie do państw o słabiej rozwiniętych regulacjach rynku pracy. Nisko opłacane kadry w mniej rozwiniętych państwach wykonują kluczowe zadania związane z filtrowaniem danych – jest to proces, bez którego najpopularniejsze modele językowe nie osiągnęłyby nawet w połowie tak wysokiego poziomu zaawansowania, jakim charakteryzują się dziś.

Sztuczna inteligencja nie uczy się bezpośrednio z istniejących już zasobów internetu, lecz z przykładowych danych, które muszą być przeanalizowane i oznaczone przez ludzi. Ich praca polega na oznaczaniu obiektów na obrazach, moderacji treści (oznaczanie mowy nienawiści, aktów przemocy) oraz na procesie nazywanym Reinforcement Learning from Human Feedback. Jest to kluczowa technika w trenowaniu sztucznej inteligencji, która służy do dopasowywania zachowania modeli do ludzkich oczekiwań i przyzwyczajeń.[1] Pracownicy oceniający wygenerowane teksty nie kierują się własną, niezależną opinią, lecz stosują szczegółowe, arbitralne instrukcje przygotowane przez firmy technologiczne. Wytyczne te odzwierciedlają perspektywę i wartości wąskiej grupy społecznej – młodszych, wykształconych mieszkańców zachodnich krajów demokratycznych.[2] Zlecający testują pracowników poprzez wprowadzanie ukrytych pytań kontrolnych ze z góry ustaloną

[1] „What is reinforcement learning from human feedback (RLHF)?” Dave Bergmann, IBM, 2023 https://www.ibm.com/think/topics/rlhf

[2] „Purchased Epistemic Erasure. AI Alignment, RLHF, and the Exploitation of Global Preference Labor” J.Jefferson, F. Velasco, 2026, s. 2 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=7118021

„poprawną” odpowiedzią. Jeśli ocena pracownika odbiega od oczekiwań zamawiającego, jego rating spada, co grozi brakiem zapłaty za wykonaną pracę lub utratą dostępu do kolejnych zleceń. W praktyce pracownik, który musi zrealizować wyznaczone normy ma na podjęcie decyzji około pół sekundy.[3] Pośpiech i brak dokładności to jedna z przyczyn, dla których modele językowe popełniają błędy i charakteryzują się specyficznym stylem budowania zdań. Pracownik niedokładnie lub niepoprawnie ocenia podane przykłady, a w konsekwencji gotowy model językowy generuje zmyślone fakty i fałszywe cytaty w absolutnie przekonujący sposób.

W 2019 roku Meta otworzyła w Nairobi hub moderacji treści zarządzany przez podwykonawcę Sama. Firma – z siedzibą w San Francisco, operująca jednak w większości w Afryce Południowej – promowała się jako podmiot realizujący misję „etycznej sztucznej inteligencji” i wyciągający ludzi z ubóstwa poprzez oferowanie im pracy w sektorze technologicznym.[4] Charakter zlecanej przez nich pracy przynosił jednak inne skutki. Ilość drastycznych treści wywoływała u pracowników ciężki zespół stresu pourazowego, bezsenność i szkody w życiu osobistym. Były moderator Daniel Motaung pozwał Meta i Sama po tym, jak nabawił się PTSD i został zwolniony za próbę organizacji związkowej. W 2023 roku kenijski Sąd Pracy, a we wrześniu 2024 roku Sąd Apelacyjny orzekły, że Meta może odpowiadać przed sądem w Kenii, mimo braku zarejestrowanego biura w tym kraju. Jest to pierwszy tego typu wyrok na świecie, przełamujący bezkarność wynikającą z zasłaniania się łańcuchem podwykonawców.[5] W 2021 roku Sama nawiązała współpracę z firmą OpenAI, znaną najbardziej jako twórcy chatbota ChatGPT, powierzając kenijskim pracownikom zadanie oznaczania najgroźniejszych materiałów z sieci (m.in. tortur, mowy nienawiści, opisów samobójstw i wykorzystywania dzieci) tak, aby ich model językowy mógł nauczyć się ich odrzucania i nie tworzyć podobnych treści, gdy zostanie już upubliczniony.[6] OpenAI płaciło firmie Sama około 12,50 USD za godzinę pracy jednego pracownika, podczas gdy sami pracownicy w Kenii otrzymywali wynagrodzenie w wysokości od 1,32 do 2 dolarów za godzinę[7]. Współpraca z OpenAI przy filtrowaniu drastycznych i toksycznych materiałów dla ChatGPT dobiegła końca po śledztwach dziennikarskich ujawniających

[3] ,,AI Data Labeling Exploitation: How *Underpaid Workers* in Kenya and the Philippines Undermine Model Safety” J. Smith, Silent Room Journal, 2026
https://silentroom.media/the-machine/ai-data-labeling-exploitation-who-makes-ai-safe
[4] tamże
[5] tamże
[6] tamże
[7] tamże

wyzysk, traumę psychiczną oraz rażącą dysproporcję między stawkami płaconymi przez OpenAI a realnym wynagrodzeniem pracowników. a początku 2023 roku Sama ogłosiła, że całkowicie wycofuje się z świadczenia usług moderacji treści, aby skoncentrować swoje działania wyłącznie na obszarze wizji komputerowej.[8]

Śledztwo dziennikarskie przeprowadzone przez The Bureau of Investigative Journalism w 2024 roku ujawniło, że cyfrowi pracownicy z krajów Globalnego Południa byli nieświadomie wykorzystywani do budowania systemów sztucznej inteligencji dla rosyjskiego aparatu inwigilacji, które miały być wykorzystywane do rozpoznawania twarzy[9]. Pracownicy mieli za zadanie obrysowywanie ludzkich sylwetek na nagraniach z kamer monitoringu oraz kategoryzowanie zdjęć, mieli także dodatkowo opcję wysyłania swoich własnych zdjęć w celu rozwijania działania algorytmów na różnych grupach etnicznych. Te dane były później wykorzystane w tworzeniu systemu sieci ponad 178 tysięcy kamer miejskich w Moskwie. System ten jest rozwijany przez firmy technologiczne, takie jak NTechLab i Tevian[10] (objęte później sankcjami unijnymi) były zlecane za pośrednictwem platformy crowdsourcingowej Toloka. Powstała ona jako część rosyjskiego giganta technologicznego Yandex i działała poprzez podmioty zarejestrowane m.in. w Holandii, Szwajcarii oraz Rosji.[11] Zlecenia trafiały do osób w skrajnie trudnej sytuacji życiowej i materialnej, m.in. w obozach dla uchodźców, gdzie dostęp do komputera zapewniany przez organizacje pomocowe stanowił często jedyną szansę na zarobek. Wykonawcy nie byli poinformowani, do czego zostanie wykorzystana praca, którą wykonywali.[12] Wyszkolony w ten sposób system posłużył rosyjskim służbom do identyfikacji oraz zatrzymań prewencyjnych osób uczestniczących w protestach opozycyjnych i antywojennych. Po protestach popierających Aleksieja Nawalnego w 2021 roku zatrzymano 454 osoby, a po jego śmierci w 2024 roku kamery posłużyły do

---

[8] ,,AI Data Labeling Exploitation: How *Underpaid Workers* in Kenya and the Philippines Undermine Model Safety" J. Smith, Silent Room Journal, 2026
https://silentroom.media/the-machine/ai-data-labeling-exploitation-who-makes-ai-safe
[9] ,,Tips to investigate AI labor abuses in the global south", Rowan Philp, iMEdD, 2026
https://lab.imedd.org/en/tips-to-investigate-ai-labor-abuses-in-the-global-south/
[10] ,,Online gig work is feeding Russia's surveillance machine", N. McIntyre , L. Kijewski , H. Munzinger, C. Huppertz, L. Kotkamp; The Bureau of Investigative Journalism, 2024
https://www.thebureauinvestigates.com/stories/2024-03-27/online-gig-work-is-feeding-russias-surveillance-machine
[11] tamże
[12] ,,Tips to investigate AI labor abuses in the global south", Rowan Philp, iMEdD, 2026
https://lab.imedd.org/en/tips-to-investigate-ai-labor-abuses-in-the-global-south/

zidentyfikowania i aresztowania co najmniej 19 osób składających uczestniczących w jego pogrzebie.[13]

Fasada autonomicznej, sztucznej inteligencji kryje za sobą masową, niewidzialną pracę ludzką skupioną w krajach Globalnego Południa. Współczesne giganty technologiczne zbudowały sukces na wykorzystywaniu taniej siły roboczej, delegując najtrudniejsze i najbardziej szkodliwe zadania do państw o niskich kosztach pracy i słabych regulacjach. Oszczędności i zaniedbania na etapie filtrowania danych obracają się nie tylko przeciwko pracownikom, ale także przeciwko użytkownikom tych technologii: prowadzą do halucynacji modeli językowych sztucznej inteligencji, luk w filtrach bezpieczeństwa oraz pogłębiania uprzedzeń algorytmicznych. Bezpieczeństwo sztucznej inteligencji, a także społeczeństwa może być przez wielu uznawane za iluzję, dopóki firmy nie przestaną przekładać zysku i walki na globalnym rynku ponad godne warunki pracy, a mocarstwa wykorzystywać tych nierówności do wspomagania aparatu opresji.

BIBLIOGRAFIA

1. ‚What is reinforcement learning from human feedback (RLHF)?" Dave Bergmann, IBM, 2023
   https://www.ibm.com/think/topics/rlhf
2. „Purchased Epistemic Erasure. AI Alignment, RLHF, and the Exploitation of Global Preference Labor" J.Jefferson, F. Velasco, 2026, s. 2
   https://papers.ssrn.com/sol3/papers.cfm?abstract_id=7118021
3. ‚AI Data Labeling Exploitation: How *Underpaid Workers* in Kenya and the Philippines Undermine Model Safety" J. Smith, Silent Room Journal, 2026
   https://silentroom.media/the-machine/ai-data-labeling-exploitation-who-makes-ai-safe
4. „Online gig work is feeding Russia's surveillance machine", N. McIntyre , L. Kijewski , H. Munzinger, C. Huppertz, L. Kotkamp; The Bureau of Investigative Journalism, 2024
   https://www.thebureauinvestigates.com/stories/2024-03-27/online-gig-work-is-feeding-russias-surveillance-machine
5. „Tips to investigate AI labor abuses in the global south", Rowan Philp, iMEdD, 2026
   https://lab.imedd.org/en/tips-to-investigate-ai-labor-abuses-in-the-global-south/

[13] „Online gig work is feeding Russia's surveillance machine", N. McIntyre , L. Kijewski , H. Munzinger, C. Huppertz, L. Kotkamp; The Bureau of Investigative Journalism, 2024
https://www.thebureauinvestigates.com/stories/2024-03-27/online-gig-work-is-feeding-russias-surveillance-machine