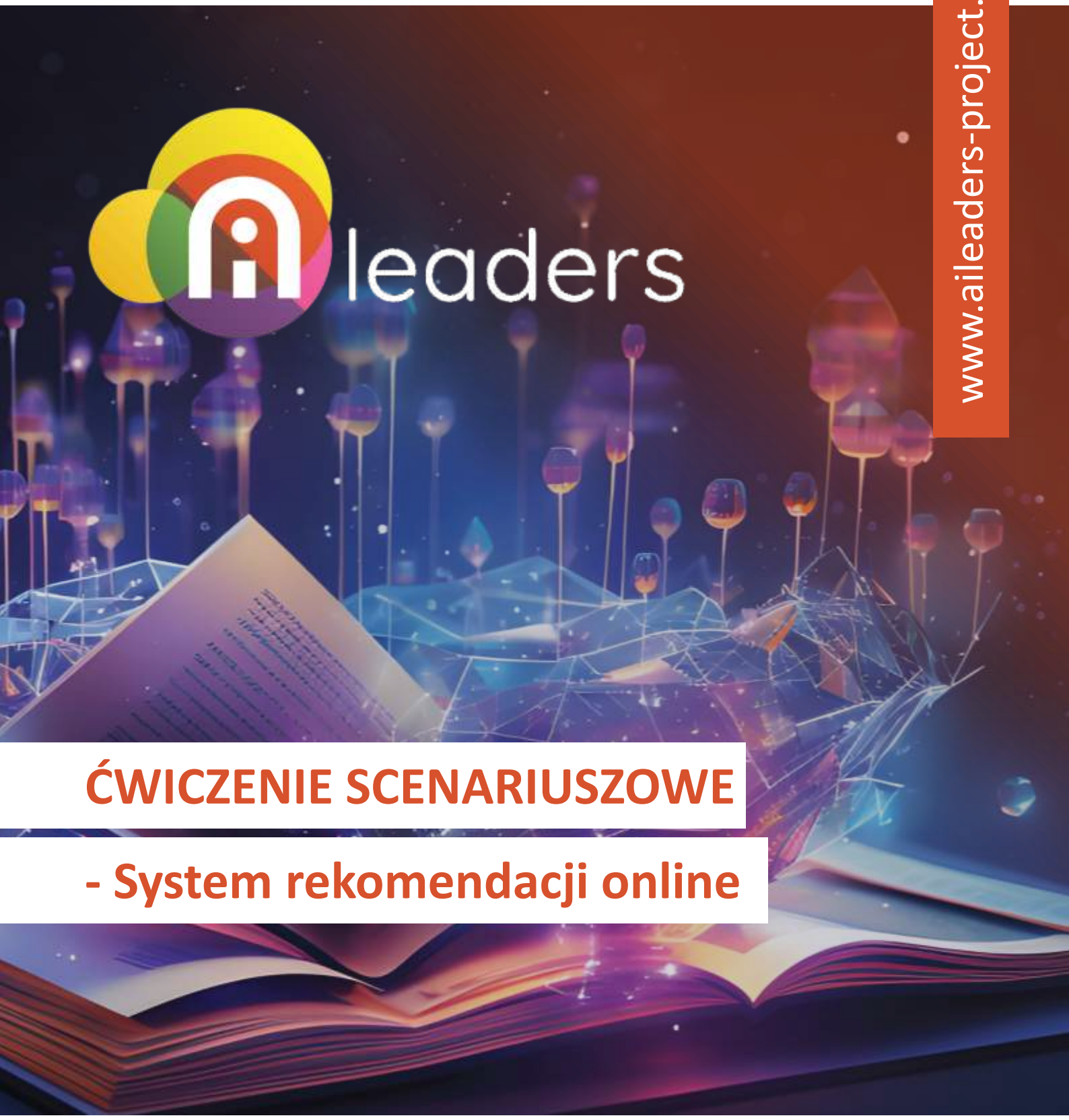
www.aileaders-project.eu

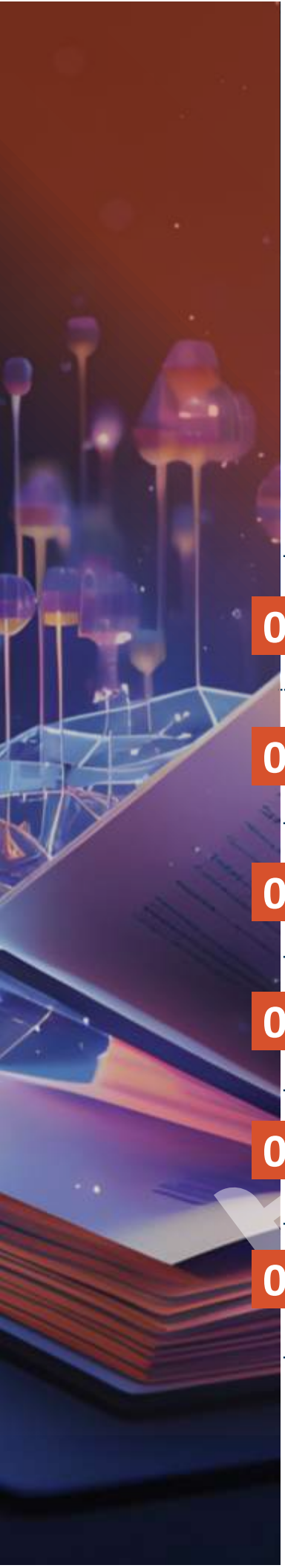


ĆWICZENIE SCENARIUSZOWE

- System rekomendacji online

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



ĆWICZENIE SCENARIUSZOWE - System rekomendacji online

01 Streszczenie

Rodzaj OER

Ćwiczenie scenariuszowe z wykorzystaniem Google Colab (riverML)

Cel/przeznaczenie

To ćwiczenie oparte na scenariuszu stawia studentów w roli praktyka uczenia maszynowego (ML), którego zadaniem jest analiza sprawiedliwości internetowego systemu rekomendacji. Bada ono, w jaki sposób ciągłe dostosowywanie się do zachowań użytkowników może z czasem wzmacniać istniejące uprzedzenia. Studenci będą mieli za zadanie zidentyfikować uprzedzenia i ocenić strategie ich łagodzenia.

Oczekiwane efekty kształcenia

Student będzie potrafił:

- Zidentyfikować potencjalne źródła stronniczości w rekomendacjach generowanych przez ML;
- Porównać wskaźniki sprawiedliwości, takie jak parytet demograficzny i równość ekspozycji;
- Zrozumieć, w jaki sposób pętle sprzężenia zwrotnego wpływają na sprawiedliwość w systemach rekomendacji online w miarę upływu czasu.

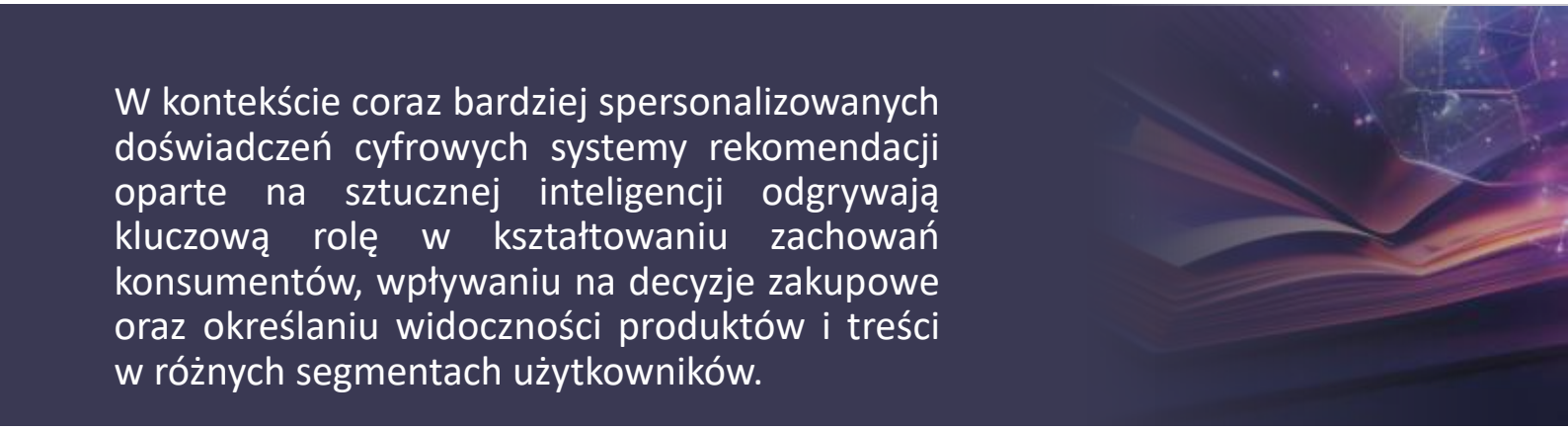
Słowa kluczowe

- Uczenie maszynowe
- Rekomendacje
- Stronniczość
- Sprawiedliwość

Sugerowane Metodologiczne Podejście

Nauka oparta na problemach.

02 Wprowadzenie



W kontekście coraz bardziej spersonalizowanych doświadczeń cyfrowych systemy rekomendacji oparte na sztucznej inteligencji odgrywają kluczową rolę w kształtowaniu zachowań konsumentów, wpływaniu na decyzje zakupowe oraz określaniu widoczności produktów i treści w różnych segmentach użytkowników.

Chociaż systemy te oferują znaczne korzyści w zakresie marketingu i optymalizacji działalności, budzą one również pilne obawy etyczne i regulacyjne, szczególnie w odniesieniu do sprawiedliwości, przejrzystości i potencjalnej dyskryminacji (Akter et al., 2022; Bozdag, 2013).

Systemy rekomendacyjne nie są jedynie artefaktami algorytmicznymi, ale złożonymi konstrukcjami socjotechnicznymi, na które wpływają decyzje ludzkie na każdym etapie – od gromadzenia danych i projektowania modeli po interpretację wyników i wdrażanie (Bozdag, 2013). Stronniczość może wynikać z wielu źródeł: ludzkich uprzedzeń, ograniczeń technicznych lub niezgodności kontekstowych. Szczególnie w środowiskach marketingowych stronniczość ta może się wzajemnie oddziaływać i nasilać, kształtując nie tylko wydajność systemu, ale także doświadczenia użytkowników i skutki społeczne (Akter et al., 2022).

Istotną kwestią jest iteracyjny charakter systemów rekomendacyjnych, które opierają się na opiniach użytkowników w celu ponownego szkolenia i udoskonalania modeli predykcyjnych. Ta pętla informacji zwrotnej może wzmacniać istniejącą nierównowagę między danymi wejściowymi a wynikami, prowadząc do utrzymujących się dysproporcji w czasie (Sun et al., 2020). Ponieważ modele stale uwzględniają nowe dane i preferencje użytkowników, ich sprawiedliwość i dokładność podlegają dynamicznym zmianom. Jeśli nie zostanie to skontrolowane, może to skutkować nadmiernym dostosowaniem do dominujących zachowań użytkowników, pogłębiając niedostateczną reprezentację grup mniejszościowych i ograniczając różnorodność treści — zjawisko związane z algorytmicznym gatekeepingiem i wzrostem popularności baniek filtrujących (Harambam et al., 2018).

Zainspirowane tymi wyzwaniami ćwiczenie oparte na scenariuszach zachęca studentów do zgłębiania etycznych aspektów systemów rekomendacyjnych, wykorzystując **zbiór danych MovieLens 100K** jako platformę testową.

Poprzez praktyczne eksperymenty studenci badają, w jaki sposób algorytmiczne pętle sprzężenia zwrotnego mogą nieumyślnie wzmacniać nierówności, w szczególności poprzez zniekształcanie widoczności treści i faworyzowanie preferencji większości. Zachęca się ich do zbadania zarówno technicznych, jak i etycznych aspektów tych systemów, łącząc wiedzę teoretyczną z praktyczną analizą. Takie podejście pedagogiczne sprzyja krytycznej świadomości ryzyka związanego z nieetyczną personalizacją, w tym utratą zaufania konsumentów, szkodą dla reputacji i ograniczonym dostępem do możliwości dla użytkowników marginalizowanych. Ćwiczenie podkreśla, że sprawiedliwość w sztucznej inteligencji nie jest statyczną oceną, ale dynamiczną, ciągłą odpowiedzialnością. Wraz z ewolucją preferencji użytkowników i zmianami treści, modele muszą się dostosowywać, aby zachować zrównoważoną wydajność we wszystkich grupach użytkowników. Ostatecznie to doświadczenie edukacyjne wyposaża studentów w narzędzia do projektowania systemów sztucznej inteligencji, które są nie tylko skuteczne i wydajne, ale także sprawiedliwe, przejrzyste i społecznie odpowiedzialne.

Wskaźniki sprawiedliwości

01 **Błąd bezwzględny oszacowania** — mierzy bezwzględny błąd między przewidywaną oceną a rzeczywistą oceną przyznaną przez użytkownika u pozycji I, definiowany jako:

$$\varepsilon_a(u,i) = |\hat{r}_{u,i} - r_{u,i}|$$

02 **Błąd przeszacowania** – mierzy, o ile prognoza przeszacowuje rzeczywistą ocenę, definiowany jako:

$$\varepsilon_o(u,i) = \max(\hat{r}_{u,i} - r_{u,i}, 0)$$

03 **Błąd niedoszacowania** – mierzy, o ile prognoza zaniża rzeczywistą ocenę, definiowany jako:

$$\varepsilon_u(u,i) = \max(r_{u,i} - \hat{r}_{u,i}, 0)$$

04 **Błąd oszacowania wartości** – mierzy błąd ze znakiem między przewidywaną oceną a rzeczywis oceną, definiowany jako:

$$\varepsilon_v(u,i) = \hat{r}_{u,i} - r_{u,i}$$

03 Prezentacja narzędzi

W tym ćwiczeniu wykorzystano **zbiór danych MovieLens 100K**, powszechnie stosowany punkt odniesienia w badaniach nad systemami rekomendacji. Zbiór danych zawiera 100 000 ocen, 943 użytkowników, 1682 filmy, a także dane demograficzne użytkowników (np. wiek, płeć, zawód).

Chociaż jest to przydatne do testowania algorytmów

Zbiór danych zawiera istotne wyzwania związane z uczciwością

01 **Nierównomierny rozkład reprezentacji w grupach demograficznych;**

02 **Nierównomierny rozkład ocen;**

03 **Nadmierna reprezentacja popularnych pozycji.**

Cechy te sprawiają, że MovieLens 100K idealnie nadaje się do badania stronniczości i sprawiedliwości algorytmów rekomendacji.

RiverML[1] to biblioteka open source w języku Python, zaprojektowana specjalnie do uczenia maszynowego na strumieniach danych. Zapewnia narzędzia do tworzenia, szkolenia i oceny modeli, które uczą się stopniowo na podstawie danych przychodzących w czasie rzeczywistym, a nie w partiach. RiverML obsługuje zarówno algorytmy uczenia nadzorowanego, jak i nienadzorowanego, oferując funkcje takie jak adaptacyjne uczenie się, wykrywanie dryfu koncepcyjnego i wstępne przetwarzanie strumieni. Jego modułowa konstrukcja ułatwia łączenie modeli, etapów wstępnego przetwarzania i strategii oceny, co czyni go potężnym narzędziem do zastosowań związanych z uczeniem się online, takich jak wykrywanie oszustw, systemy rekomendacji i analiza danych z czujników. Metryki sprawiedliwości dostępne w bibliotece zostały wdrożone przez zespół FEP, przyczyniając się do etycznej oceny modeli strumieniowych.

Narzędzie to umożliwia wykonywanie procesów oceny sprawiedliwości oraz wizualne porównanie skuteczności rekomendacji w różnych grupach i w czasie.

1. Samouczek Riverml — https://riverml.xyz/latest/introduction/installation/

04 Ćwiczenia praktyczne

Ćwiczenie

01 Wejdź na stronę https://tinyurl.com/5n8vvpx9, żeby otworzyć dostarczony notatnik Jupyter.

02 Uruchom wszystkie komórki w notatniku (kliknij kartę Runtime, a następnie Run All).

03 Przeanalizuj i porównaj wskaźniki sprawiedliwości w różnych grupach użytkowników (np. według płci, wieku).

04 Obserwuj, jak zmieniają się wskaźniki sprawiedliwości w miarę gromadzenia większej liczby interakcji użytkowników.

Pytania do dyskusji

01 Które grupy użytkowników są systematycznie faworyzowane lub dyskryminowane?

02 Jak zmienia się jakość rekomendacji w czasie?

03 Jakie kompromisy pojawiają się przy próbie zrównoważenia zaangażowania i sprawiedliwości?

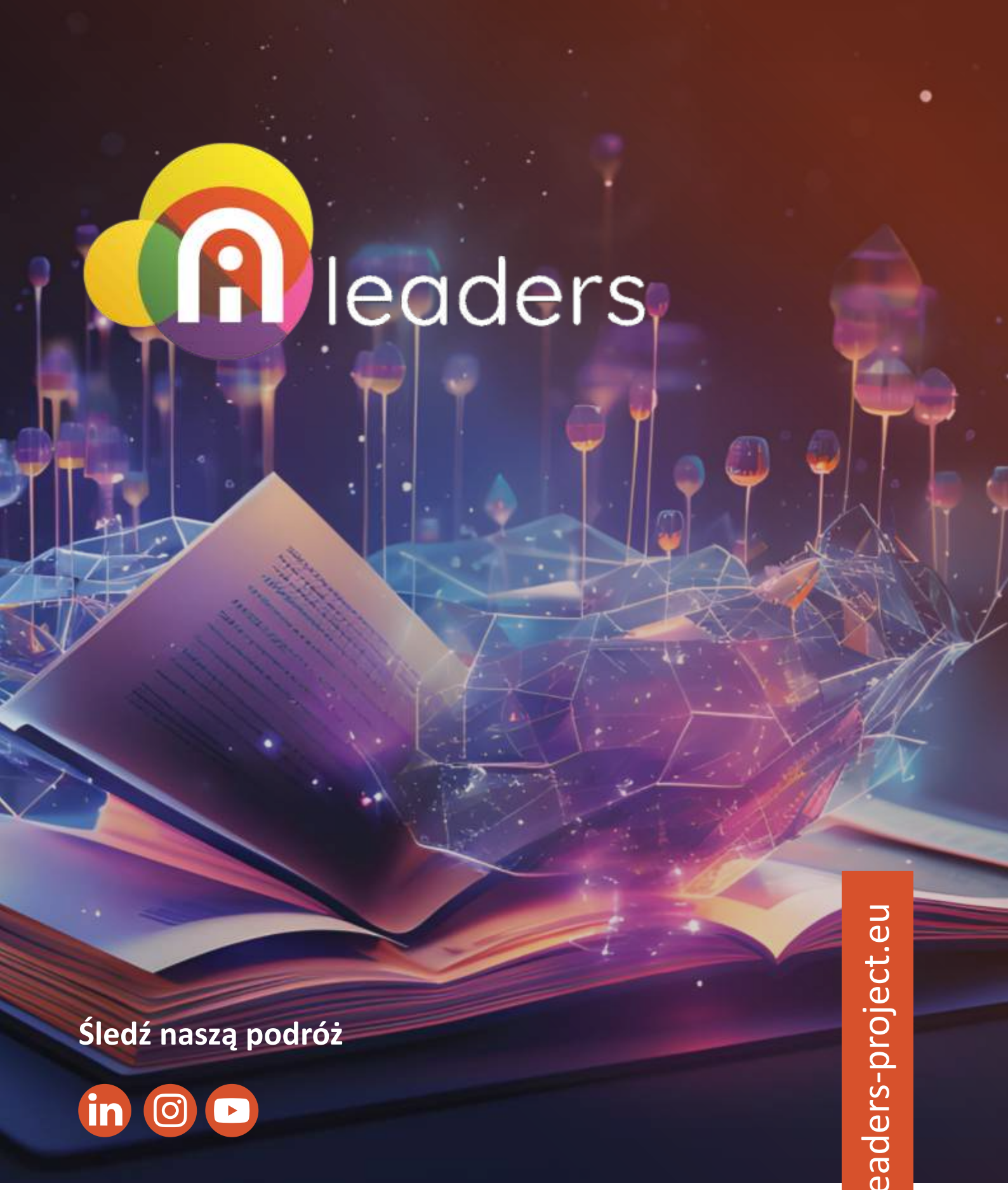
05 Wnioski

To ćwiczenie ilustruje dynamiczny charakter sprawiedliwości w internetowych systemach rekomendacji.

Chociaż początkowa wydajność systemu może wydawać się obiektywna, długotrwałe działanie może prowadzić do znacznych dysproporcji spowodowanych pętlami sprzężenia zwrotnego i stronniczością popularności. Te nierówności wpływają na doświadczenia użytkowników i ich zaufanie, szczególnie w przypadku grup użytkowników niedostatecznie reprezentowanych lub mniejszościowych. Stosując wskaźniki sprawiedliwości i analizując ewolucję stronniczości w czasie rzeczywistym, studenci uzyskują praktyczną wiedzę na temat etycznego wdrażania uczenia maszynowego. Badają również ograniczenia statycznej oceny i znaczenie ciągłego monitorowania sprawiedliwości w systemach adaptacyjnych. Ćwiczenie podkreśla napięcie między optymalizacją zaangażowania a zapewnieniem sprawiedliwego traktowania – kluczowym wyzwaniem dla etycznego projektowania sztucznej inteligencji.

06 Referencje

- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. Ethics and Information Technology, 15(3), 209–227.
- Akter, S., Michael, K., Bandara, R., Wamba, S.F., Foropon, C., & Papadopoulos, T. (2022). Algorithmic bias in machine learning–based marketing models: A dynamic capability view. Journal of Business Research.
- Harambam, J., Helberger, N., & van Hoboken, J. (2018). Democratizing algorithmic news recommenders: How to materialize voice in a technologically saturated media ecosystem. Big Data & Society.
- Sun, T., Gautam, K., & Joachims, T. (2020). Debiasing Learning from User Interaction. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.



Śledź naszą podróż

www.aileaders-project.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.