Wykonali:
Bartosz Bartocha
Filip Jastrzębski
Jakub Pluczak
Piotr Słomka
Władysław Waśkiewicz

TruthLens

01 Wprowadzenie do TruthLens

Koncepcja projektu

Wykrywanie manipulacji w mediach

TruthLens wykrywa manipulacje mediów poprzez analizę niespójności, deepfake'ów i zmienionych treści w wiadomościach i artykułach, zapewniając wiarygodną weryfikację informacji.

Automatyczna weryfikacja źródeł

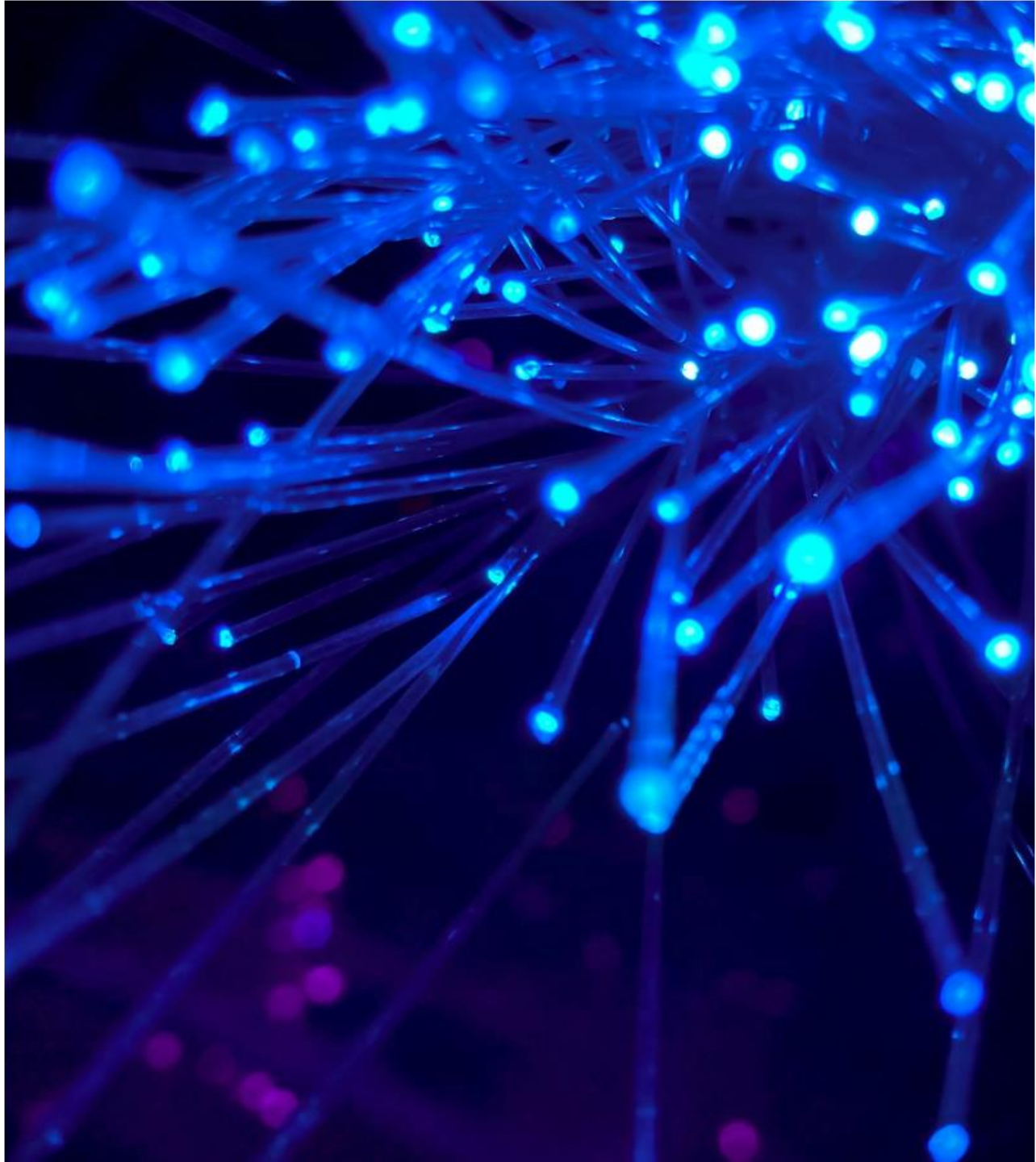
TruthLens weryfikuje źródła wiadomości i artykułów poprzez porównywanie danych, wykrywanie niespójności i ocenę wiarygodności, aby zapewnić dokładną i rzetelną analizę informacji.

Zaznaczanie podejrzanych treści

TruthLens wykrywa i oznacza podejrzane treści poprzez analizę ich autentyczności, pomagając użytkownikom szybko i skutecznie identyfikować potencjalne fałszywe informacje.

Udostępnianie bezpieczniejszych wersji treści

TruthLens automatycznie wykrywa fałszywe informacje, umożliwiając użytkownikom udostępnianie zweryfikowanych, bezpieczniejszych wersji wiadomości i artykułów, promując dokładne i wiarygodne treści w Internecie.



Znaczenie projektu

Rosnący problem dezinformacji

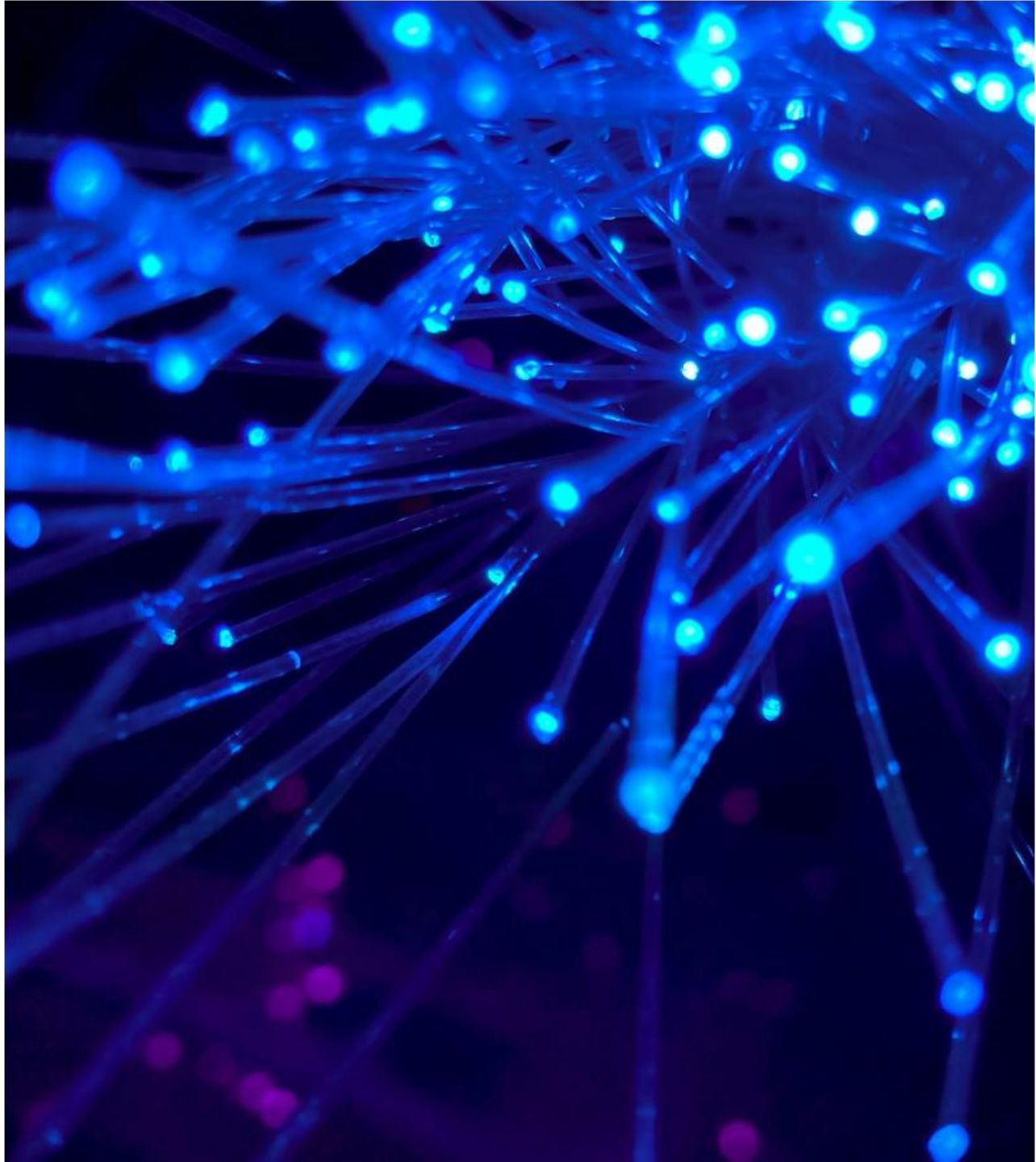
Szybkie rozprzestrzenianie się dezinformacji na platformach cyfrowych powoduje zamieszanie, nieufność i podziały społeczne, co podkreśla pilną potrzebę stworzenia niezawodnych narzędzi do automatycznej weryfikacji autentyczności wiadomości i mediów.

Potrzeba zautomatyzowanych narzędzi

Gwałtowny wzrost liczby fałszywych informacji wymaga zastosowania zautomatyzowanych narzędzi, takich jak TruthLens, które pozwalają szybko zweryfikować dokładność treści, zapewniając podejmowanie świadomych decyzji i skutecznie zwalczając rozprzestrzenianie się fałszywych wiadomości.

Korzyści dla użytkowników i społeczeństwa

Zwiększa kompetencje medialne, zwalcza dezinformację i wzmacnia pozycję użytkowników dzięki rzetelnym analizom, sprzyjając podejmowaniu świadomych decyzji i promując godne zaufania środowisko informacji cyfrowej dla społeczeństwa.



02 Zrozumienie dezinformacji i manipulacji

Rodzaje dezinformacji

Fałszywe wiadomości i mistyfikacje

Fałszywe wiadomości i mistyfikacje to celowo wprowadzające w błąd informacje, które często szybko rozprzestrzeniają się w mediach społecznościowych, powodując zamieszanie i podważając zaufanie publiczne.

Deepfake i zmodyfikowane media

Deepfake i zmodyfikowane media manipulują treściami audiowizualnymi w celu oszukania odbiorców, utrudniając odróżnienie rzeczywistości od fałszu, co zwiększa ryzyko dezinformacji na platformach cyfrowych.

Dezinformacje w mediach społecznościowych

Dezinformacja rozprzestrzenia się szybko w mediach społecznościowych poprzez emocjonalne, wprowadzające w błąd treści mające na celu manipulowanie opiniami, często przy użyciu fałszywych kont i algorytmów w celu wzmocnienia fałszywych narracji.

Społeczny wpływ fałszywych treści

Destabilizacja polityczna

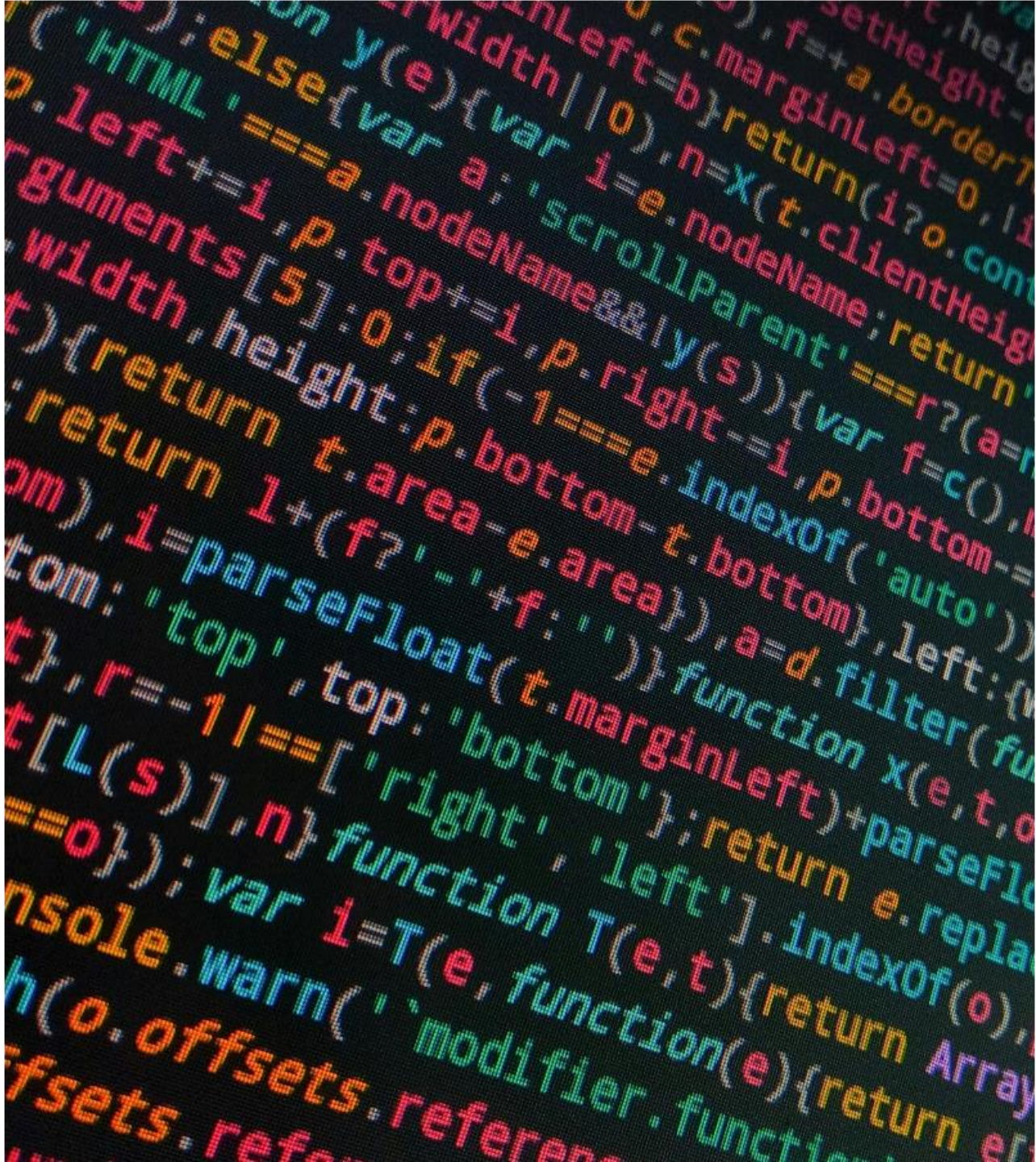
Dezinformacja podsyca niepokoje polityczne poprzez rozpowszechnianie fałszywych narracji, podważanie zaufania do instytucji i polaryzację społeczności, co ostatecznie destabilizuje procesy demokratyczne i osłabia spójność społeczną.

Panika związana ze zdrowiem publicznym

Fałszywe twierdzenia dotyczące zdrowia, zwłaszcza dezinformacja dotycząca szczepień, podsycają strach społeczny, zmniejszają wskaźniki szczepień i utrudniają walkę z pandemią, zwiększając szkody społeczne i nieufność wobec władz medycznych.

Konflikty społeczne i polaryzacja

Dezinformacja pogłębia konflikty społeczne poprzez rozpowszechnianie fałszywych narracji, pogłębianie polaryzacji i podważanie zaufania do instytucji, co prowadzi do fragmentacji społeczności i pogłębienia podziałów społecznych.



03 Podejście techniczne TruthLens

Zautomatyzowana analiza treści

Algorytm wykrywania manipulacji

Wykorzystuje zaawansowane uczenie maszynowe do wykrywania niespójności wizualnych, porównywania metadanych i identyfikowania deepfake'ów, umożliwiając dokładne wykrywanie manipulacji na zdjęciach i w treściach mediów społecznościowych.

Zgłaszanie i sygnalizowanie podejrzanych treści

Wykorzystuje sztuczną inteligencję do wykrywania niespójności, deepfake'ów i dezinformacji poprzez analizę danych wizualnych i tekstowych, umożliwiając flagowanie w czasie rzeczywistym i szczegółowe raportowanie podejrzanych treści.

Techniki weryfikacji źródła

Wykorzystuje deep learning i krzyżowe odniesienia metadanych w celu weryfikacji autentyczności źródła, sprawdzania historii publikacji i wykrywania niespójności, zapewniając niezawodną identyfikację prawdziwych i zmanipulowanych wiadomości oraz treści medialnych.

04 Doświadczenie użytkownika i interfejs

Jak użytkownicy korzystają z aplikacji

Przesyłanie i analiza treści

Użytkownicy przesyłają artykuły lub zdjęcia, a aplikacja automatycznie analizuje i wyświetla wyniki, podkreślając autentyczność, kluczowe informacje i potencjalne nieprawdziwe informacje, aby ułatwić przeglądanie.

Otrzymywanie ostrzeżeń i zaleceń

Użytkownicy otrzymują jasne, terminowe ostrzeżenia i dostosowane do ich potrzeb rekomendacje oparte na wiarygodności treści, co ułatwia podejmowanie świadomych decyzji i promuje bezpieczniejsze i bardziej wiarygodne korzystanie z mediów w ramach aplikacji.

Cechy edukacyjne

Wyjaśnienie wykrytych manipulacji

Interfejs podkreśla wykryte manipulacje za pomocą wyraźnych znaczników wizualnych i zwięzłych wyjaśnień, pomagając użytkownikom intuicyjnie i edukacyjnie zrozumieć charakter i wpływ zmienionych treści.

Zapewnianie wyników wiarygodności źródła

Integruje oceny wiarygodności źródeł w czasie rzeczywistym, pomagając użytkownikom szybko ocenić wiarygodność wiadomości i filmów, ułatwiając podejmowanie świadomych decyzji dzięki przejrzystym, łatwym do zrozumienia wskaźnikom zaufania w interfejsie.

Wskazówki dotyczące wykrywania fałszywych treści

Szukaj niespójnych szczegółów, weryfikuj źródła, sprawdzaj, czy nie zmieniono elementów wizualnych, i zachowaj ostrożność w przypadku sensacyjnych nagłówków. Porównuj informacje z wielu wiarygodnych źródeł, aby skutecznie wykrywać fałszywe treści.

Wyzwania i 05 kierunki rozwoju na przyszłość

Ograniczenia obecnych modeli sztucznej inteligencji

Dokładność i wyniki fałszywie dodatnie

Obecne modele sztucznej inteligencji borykają się z wyzwaniami związanymi z równoważeniem dokładności i minimalizowaniem liczby fałszywych alarmów, często błędnie klasyfikując treści o niuansach, co ogranicza wiarygodność i zaufanie użytkowników do automatycznej analizy wiadomości i filmów.

Ewolucja technik manipulacji

Szybko ewoluujące techniki manipulacji, takie jak postępy w dziedzinie deepfake i zniekształcanie kontekstu, ograniczają zdolność sztucznej inteligencji do dokładnego wykrywania wyrafinowanych i nowych wzorców dezinformacji.

Kwestie dotyczące prywatności i etyki

Obecne systemy sztucznej inteligencji narażają prywatność twarzy na ryzyko związane z gromadzeniem danych i potencjalnymi uprzedzeniami, budząc obawy etyczne dotyczące nadzoru, zgody i sprawiedliwości w automatycznej analizie wiadomości i filmów.

Plan działania w przyszłości

1

Ulepszanie algorytmów wykrywania

Poprawa dokładności algorytmów poprzez deep learning, wykorzystanie fuzji danych multimodalnych oraz ciągłą aktualizację modeli w celu dostosowania się do zmieniających się taktyk dezinformacyjnych i różnorodnych formatów treści.

2

Rozszerzenie zasięgu platformy

Rozszerzenie możliwości sztucznej inteligencji w celu analizowania różnorodnych platform wykraczających poza obecny zakres, w tym nowych mediów społecznościowych i sieci regionalnych, zapewniając kompleksowe wykrywanie w różnych formatach treści i zachowaniach użytkowników.

3

Współpraca z organizacjami zajmującymi się weryfikacją faktów

Zwiększ dokładność poprzez współpracę z organizacjami zajmującymi się weryfikacją faktów, aby uzyskać dostęp do sprawdzonych danych, usprawnić proces walidacji w czasie rzeczywistym oraz budować zaufanie poprzez wspólne aktualizacje i dzielenie się wiedzą specjalistyczną w zakresie wykrywania dezinformacji.

06 Wnioski

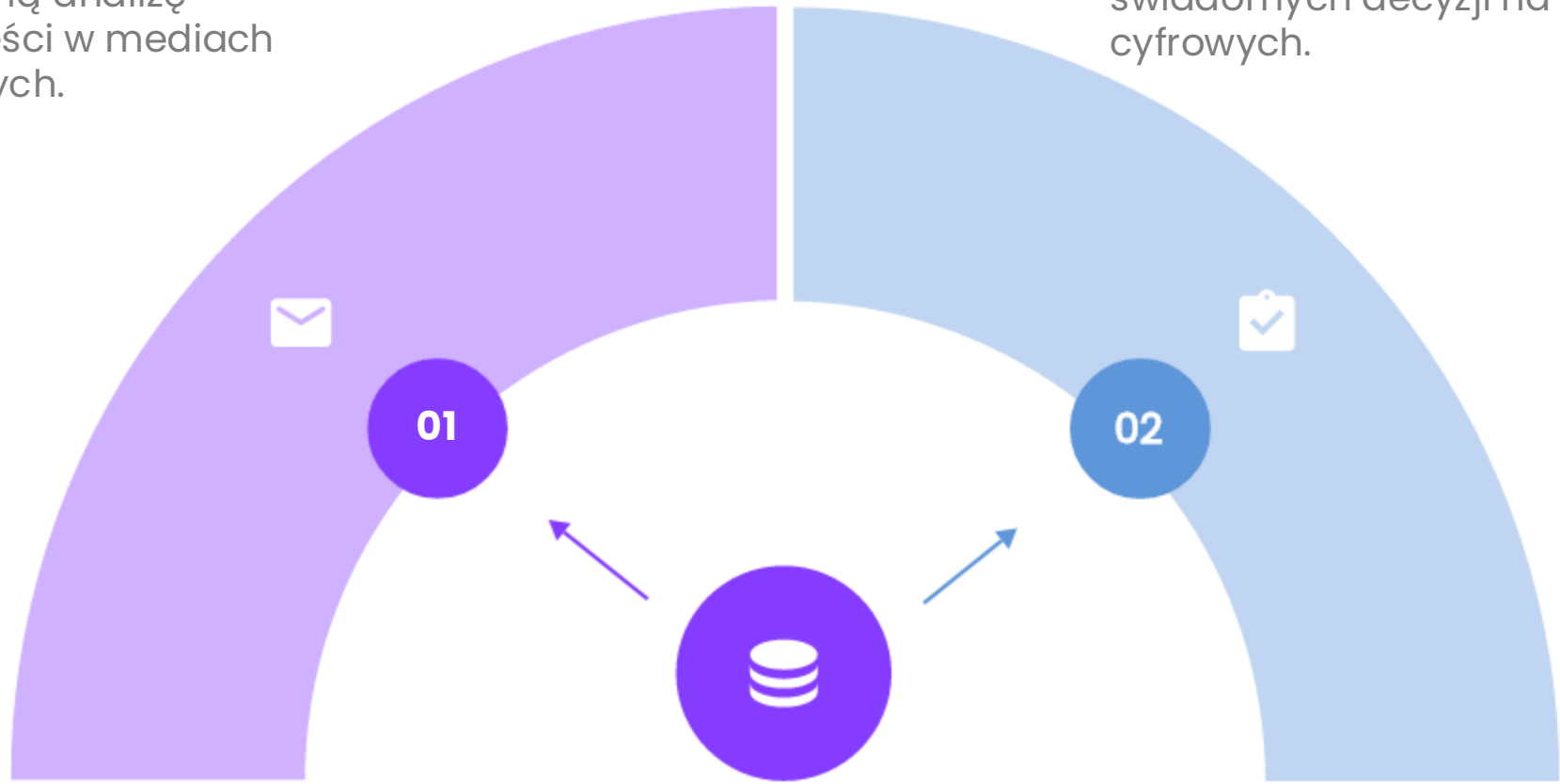
Podsumowanie zalet TruthLens

Skuteczna walka z dezinformacją

TruthLens umożliwia użytkownikom szybkie wykrywanie i zwalczanie dezinformacji, zwiększając świadomość medialną i budując zaufanie poprzez rzetelną, zautomatyzowaną analizę wiadomości i treści w mediach społecznościowych.

Wspieranie użytkowników w podejmowaniu świadomych decyzji

Umożliwia użytkownikom natychmiastową weryfikację autentyczności treści, promując krytyczne myślenie i zapobiegając dezinformacji, a ostatecznie budując zaufanie i sprzyjając podejmowaniu świadomych decyzji na platformach cyfrowych.



Zachęcanie do zaangażowania społecznego

Zaangażowanie młodzieży i edukacja

Wzmacnia pozycję młodzieży poprzez rozwój umiejętności cyfrowych, promując krytyczne myślenie i odpowiedzialne dzielenie się informacjami, aby budować czujną, świadomą społeczność, która wspólnie zwalcza dezinformację.

Wspólna odpowiedzialność za bezpieczniejsze media

Umożliwienie użytkownikom identyfikowania fałszywych informacji sprzyja zbiorowej odpowiedzialności, promując bezpieczniejsze ekosystemy medialne poprzez aktywną czujność społeczności i współpracę w zakresie weryfikacji i udostępniania dokładnych informacji.

Dziękujemy za uwagę