

SAINT Program Nauczania

UNIT 3: Zastosowanie sztucznej inteligencji w mowie i wizji

Deliverable: WP2/2.2



SAINT

HANDS ON INTRODUCTION TO ARTIFICIAL
INTELLIGENCE IN PRIMARY EDUCATION
USING MINECRAFT

KWIECIEŃ 2023

EKVASIS

Opracowany przez: EKVASIS Team

Numer projektu: 2022-1-FR01-KA220-SCH-000087794



**Co-funded by
the European Union**

Wsparcie Komisji Europejskiej dla powstania tej publikacji nie oznacza poparcia dla jej treści, które odzwierciedlają jedynie poglądy autorów, a Komisja nie ponosi odpowiedzialności za jakiegokolwiek wykorzystanie informacji w niej zawartych.

HISTORIA REWIZJI

Wersja	Data	Autor	Opis	Działanie	Strony
1.0	17/01/2023	TARAN	Utworzenie	C	TBS

(*) Działanie: U = Utworzenie, W = Wstaw, A = Aktualizuj, Z = Zastąp, U = Usuń

DOKUMENTY REFERENCYJNE

ID	Referencja	Tytuł
1	2022-1-FR01-KA220-SCH-000087794	SAINT Wniosek
2		

STOSOWNE DOKUMENTY

ID	Referencja	Tytuł
1		
2		

Spis treści

1	Wprowadzenie do projektu	4
1.1	Zakres projektu.....	4
1.2	Grupy docelowe	5
1.3	Cel niniejszego dokumentu.....	5
2	Słowniczek urzędzenia	5
3	Wprowadzenie do jednostki.....	6
3.1	Opis.....	6
3.2	Cele i wyniki nauczania	6
3.3	Szacowany czas szkolenia	7
4	Treść kursu Jednostki.....	7
4.1	Wprowadzenie.....	7
4.2	Pomysł 1: Percepcja.....	7
4.3	Pomysł 2: Reprezentacja i rozumowanie	9
4.4	Pomysł 3: Uczenie się	10
4.5	Pomysł 4: Naturalna interakcja	12
4.6	Pomysł 5: Wpływ społeczny	13
4.7	Studia przypadków i historie sukcesu	14
5	Dodatkowe materiały i zasoby	15
6	Podsumowanie.....	17
7	Quiz.....	17
8	Referencje	19

1 Wprowadzenie do projektu

1.1 Zakres projektu

Działając jako idealne cyfrowe środowisko edukacyjne do nauczania dzieci o praktycznych zastosowaniach sztucznej inteligencji w oparciu o wytyczne projektu AI4K12, motywacja tego projektu obejmuje następujące cele:

- Zapoznanie uczniów, nauczycieli i wychowawców z koncepcjami sztucznej inteligencji, jej wpływem na nasze społeczeństwo i związanymi z nią praktycznymi wdrożeniami,
- zaspokojenie rosnącej potrzeby opracowania rozwiązań zdalnego nauczania ułatwiających zaangażowanie uczniów, ich kreatywność, umiejętność rozwiązywania problemów i podejmowania decyzji,
- Podnoszenie kwalifikacji nauczycieli i edukatorów za pomocą nowych zestawów umiejętności (PBL, AI, grywalizacja itp.) opracowanych dzięki innowacyjnym metodom nauczania,
- Poprawa wskaźników zaangażowania dzieci poprzez zastosowanie innowacyjnego sposobu nauczania, pomagającego dzieciom rozwijać kreatywność,
- Zmniejszenie luki między potrzebą a dostępnością umiejętności związanych ze sztuczną inteligencją.

AI Adventures in Minecraft uczy umiejętności związanych ze sztuczną inteligencją dzieci w wieku od 9 do 12 lat, wykorzystując świat Minecraft. Dzięki temu tworzymy zabawne, interaktywne i kreatywne środowisko uczenia się poprzez konkretne działania i wyzwania zgodne z wytycznymi AI4K12 (ai4ka12.org) i 5 wielkimi ideami sztucznej inteligencji: 1) Percepcja, 2) Reprezentacja i rozumowanie, 3) Uczenie się, 4) Naturalne interakcje, 5) Wpływ społeczny.

W tym celu projekt rozwija i promuje następujące namacalne rezultaty:

- Program nauczania: kompletny kurs nauczania wprowadzający sztuczną inteligencję do nauczania szkolnego w oparciu o 5 głównych idei ram AI4K12. Kurs rozpowszechnia wiedzę na temat wytycznych edukacyjnych AI4K12 dotyczących sztucznej inteligencji i 5 wielkich idei, bada wpływ sztucznej inteligencji na nasze społeczeństwo i poprawia zrozumienie odpowiednich pojęć.
- Dostosowany świat Minecraft (AI Adventures World) dostarczający wyzwań edukacyjnych opartych na kursie edukacyjnym. Wykorzystuje on koncepcję "escape room" i oferuje działania związane z uczeniem się opartym na problemach. Jedno wyzwanie dla każdej jednostki lub lekcji.
- Wirtualna przestrzeń odlewni wspierająca rosnącą społeczność użytkowników SAINT i kierująca korekcyjną/perfekcyjną i ewolucyjną konserwacją pakietu szkoleniowego.

1.2 Grupy docelowe

Projekt zakłada bezpośrednie zaangażowanie nauczycieli, głównie nauczycieli dzieci w wieku 9-12 lat lub pracowników szkolnictwa wyższego zaangażowanych w nauczanie pedagogów. Nauczyciele ci są albo nauczycielami przedmiotów STEM, albo mają pewną wiedzę i zainteresowanie sztuczną inteligencją i/lub grą Minecraft.

Jeśli chodzi o zidentyfikowane pośrednie grupy docelowe, zaangażowane mogą być następujące osoby:

- Centra STEM chcące rozwijać swój katalog innowacyjnych technologii nauczania lub katalog produktów zwiększających wiedzę na temat AI,
- Instytucje szkolnictwa wyższego współpracujące z firmami / władzami publicznymi zaangażowanymi w tworzenie materiałów edukacyjnych,
- organizacje, stowarzyszenia lub sieci, które chcą zapewnić rodzicom i / lub nauczycielom materiały edukacyjne na temat sztucznej inteligencji: takie jak kluby kodowania, ośrodki kształcenia dorosłych, usługi coachingu przedsiębiorczości, centra kształcenia ustawicznego itp.

1.3 Cel niniejszego dokumentu

Pakiet roboczy nr 2 - Program edukacyjny AI4K12 koncentruje się na stworzeniu kompletnego kursu na temat sztucznej inteligencji wraz z zestawem 5 wyzwań w powiązonym świecie Minecraft, aby zilustrować praktyczne wdrożenie technologii.

Niniejszy program nauczania sztucznej inteligencji składa się w sumie z 5 jednostek materiałów pedagogicznych opartych na wytycznych edukacyjnych AI4K12 i celach edukacyjnych przedstawionych w następstwie badań krajowych:

1. Zastosowanie sztucznej inteligencji w uczeniu maszynowym,
2. Zastosowanie sztucznej inteligencji w pracy i przedsiębiorczości,
3. Zastosowanie sztucznej inteligencji w mowie i wizji,
4. Zastosowanie sztucznej inteligencji w grach i łamigłówkach,
5. Zastosowanie sztucznej inteligencji w życiu codziennym.

Dodatkowo w każdej jednostce utworzono glosariusz, aby ułatwić przyjęcie pakietu SAINT przez nauczycieli i szkoły.

2 Słowniczek urzędzenia

Słowa	Definicja
Uczenie maszynowe	Podzbiór sztucznej inteligencji, który wykorzystuje algorytmy i modele statystyczne, aby umożliwić systemom komputerowym poprawę ich wydajności w określonym zadaniu

Rozpoznawanie mowy	na podstawie wprowadzonych danych, bez konieczności ich wyraźnego programowania.
Wizja komputerowa	Zdolność maszyny lub programu do identyfikowania i rozumienia ludzkiej mowy.
Uczenie głębokie	często wykorzystywana w aplikacjach takich jak wirtualni asystenci, zautomatyzowana obsługa klienta i transkrypcja.
Przetwarzanie języka naturalnego (NLP)	Dziedzina nauki skupiająca się na umożliwieniu komputerom interpretowania i rozumienia danych wizualnych z otaczającego je świata, często wykorzystywana w aplikacjach takich jak rozpoznawanie obiektów, twarzy i autonomicznych pojazdów.
Uczenie ze wzmocnieniem	autonomiczne pojazdy.
Rozszerzona rzeczywistość	Podzbiór uczenia maszynowego, który wykorzystuje sztuczne sieci neuronowe do
Wirtualna rzeczywistość	modelowania złożonych wzorców w danych, często wykorzystywanych w aplikacjach takich jak rozpoznawanie obrazów i mowy.

3 Wprowadzenie do jednostki

3.1 Opis

Ten moduł wprowadzi uczestników w ekscytującą dziedzinę sztucznej inteligencji i jej zastosowania w mowie i wizji. Omówione zostaną różne rodzaje uczenia maszynowego, w tym uczenie nadzorowane, nienadzorowane i wzmacniające, a także sposób ich wykorzystania w rozpoznawaniu mowy, klasyfikacji obrazów i innych rzeczywistych zastosowaniach. Dział ten zajmie się również naturalną interakcją, na przykład tym, w jaki sposób sztuczna inteligencja wchodzi w interakcję z ludźmi oraz jak rzeczywistość wirtualna i rozszerzona zmieniają sposób, w jaki wchodzimy w interakcję z technologią. Dodatkowo, w module omówiony zostanie społeczny wpływ sztucznej inteligencji, w tym kwestie etyczne oraz zalety i wady technologii sztucznej inteligencji. Pod koniec tego rozdziału uczniowie powinni posiadać podstawową wiedzę na temat sztucznej inteligencji w mowie i wizji, a także jej potencjału w rozwiązywaniu rzeczywistych problemów.

3.2 Cele i wyniki nauczania

W tej części uczestnicy zapoznają się z zastosowaniem sztucznej inteligencji w mowie i wizji. Poznają podstawy uczenia maszynowego, w tym uczenie nadzorowane, nienadzorowane i wzmacniające. Jednostka obejmie również naturalną interakcję i jej znaczenie w sztucznej inteligencji, a także wpływ społeczny i kwestie etyczne związane z wykorzystaniem sztucznej inteligencji.

Po pomyślnym ukończeniu tego modułu uczestnicy:

- powinni być w stanie zidentyfikować różne rodzaje uczenia maszynowego i ich zastosowania w mowie i wizji.

- będą w stanie wyjaśnić znaczenie naturalnej interakcji i jej roli w rozwoju systemów sztucznej inteligencji.
- będzie miał solidne zrozumienie społecznego wpływu sztucznej inteligencji i kwestii etycznych, które należy wziąć pod uwagę przy opracowywaniu i wdrażaniu tych systemów.
- będzie w stanie ocenić rzeczywiste zastosowania sztucznej inteligencji w mowie i wizji oraz rozpoznać ich potencjał w rozwiązywaniu złożonych problemów.

3.3 Szacowany czas szkolenia

Ukończenie modułu wraz z wdrożeniem dostarczonej wiedzy potrwa około 5-6 godzin, w oparciu o ilość treści i quizów zawartych w module. Rzeczywisty czas ukończenia modułu i zastosowania zdobytej wiedzy może się jednak różnić w zależności od tempa nauki i znajomości tematu.

4 Treść kursu Jednostki

4.1 Wprowadzenie

Witamy w dziale poświęconym zastosowaniu sztucznej inteligencji w mowie i wizji. W tym dziale zbadamy, w jaki sposób sztuczna inteligencja jest wykorzystywana do przekształcania dziedzin rozpoznawania mowy i wizji komputerowej. Zaczniemy od omówienia podstaw uczenia maszynowego i różnych technik stosowanych w aplikacjach mowy i wizji, w tym uczenia nadzorowanego, nienadzorowanego i wzmacniającego. Następnie zagłębimy się w naturalną interakcję i sposób, w jaki sztuczna inteligencja zmienia sposób interakcji z maszynami, w tym interakcji mowy i wizji. Zbadamy również społeczny wpływ sztucznej inteligencji w zastosowaniach mowy i wizji, w tym jej zalety i wady oraz kwestie etyczne. Wreszcie, przedstawimy niektóre z najbardziej ekscytujących studiów przypadków i historii sukcesu w tej dziedzinie, pokazując, w jaki sposób sztuczna inteligencja jest wykorzystywana do rozwiązywania rzeczywistych problemów w opiece zdrowotnej, edukacji i rozrywce. Pod koniec tego rozdziału będziesz miał solidną wiedzę na temat tego, w jaki sposób sztuczna inteligencja przekształca dziedziny mowy i widzenia oraz jaki może mieć potencjalny wpływ na całe społeczeństwo.

4.2 Pomysł 1: Percepcja

Sztuczna inteligencja (AI) to szybko rozwijająca się dziedzina, która zmienia nasze życie na wiele sposobów. Od rozpoznawania mowy po analizę obrazu, sztuczna inteligencja stała się integralną częścią naszego codziennego życia. Jednak nauczanie dzieci pojęć związanych ze sztuczną inteligencją może być trudne, zwłaszcza biorąc pod uwagę złożony charakter tej technologii. Aby sprostać temu wyzwaniu, opracowano ramy AI4K12 jako zestaw wytycznych do nauczania koncepcji sztucznej inteligencji i praktycznych zastosowań dla uczniów w wieku K-12. Ramy te opierają się na pięciu dużych ideach, z których pierwsza to Percepcja. W tej jednostce zbadamy pierwszą dużą ideę ramy AI4K12, Percepcję, która obejmuje zrozumienie, w jaki sposób sztuczna inteligencja postrzega otaczający nas świat. Znaczenie percepcji w sztucznej inteligencji jest nie do przecenienia, ponieważ pozwala ona maszynom interpretować i reagować na otaczający je świat w sposób podobny do tego,

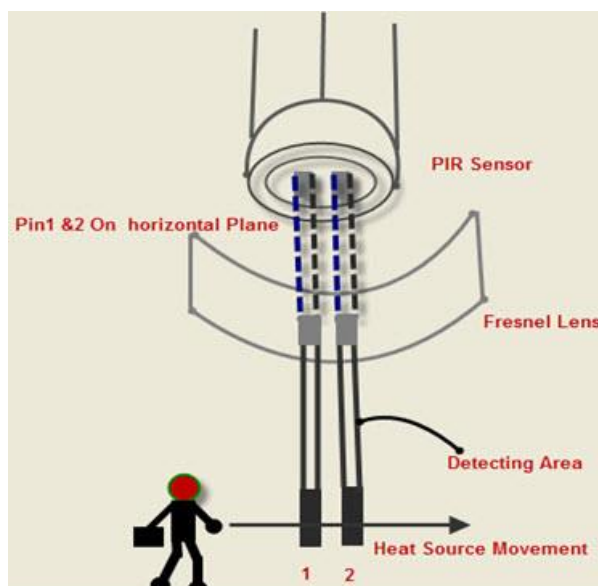
w jaki robią to ludzie. Aby postrzegać otaczający je świat, systemy AI opierają się na szeregu czujników, z których każdy został zaprojektowany do zbierania danych z różnych źródeł. Czujniki te odgrywają kluczową rolę w umożliwieniu maszynom zrozumienia ich otoczenia i podejmowania decyzji na podstawie tych informacji. W tej sekcji przedstawimy różne rodzaje czujników używanych w sztucznej inteligencji i sposób ich działania.

Przykłady:

Mikrofony: Jednym z najpopularniejszych typów czujników wykorzystywanych w sztucznej inteligencji jest mikrofon, który służy do rozpoznawania mowy. Mikrofony zbierają fale dźwiękowe i przekształcają je w sygnały cyfrowe, które mogą być przetwarzane przez algorytmy sztucznej inteligencji. Asystenci głosowi, tacy jak Alexa firmy Amazon i Siri firmy Apple, polegają na mikrofonach, aby rozumieć polecenia głosowe i reagować na nie.

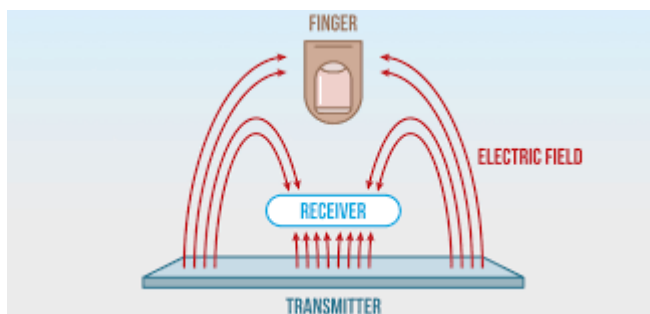
Kamery: Innym krytycznym typem czujnika wykorzystywanym w sztucznej inteligencji jest kamera, która jest używana do wizji komputerowej. Kamery przechwytyują obrazy lub klatki wideo i wykorzystują algorytmy do wydobywania z nich informacji. Informacje te mogą obejmować wykrywanie obiektów, rozpoznawanie twarzy i śledzenie ruchu. Kamery są wykorzystywane w wielu rzeczywistych zastosowaniach, w tym w systemach nadzoru, pojazdach autonomicznych i kamerach bezpieczeństwa.

Czujniki podczerwieni: Czujniki podczerwieni wykrywają promieniowanie podczerwone, które jest emitowane przez wszystkie obiekty o temperaturze powyżej zera bezwzględnego. Czujniki te są wykorzystywane w aplikacjach takich jak domowe systemy bezpieczeństwa i kamery noktowizyjne. Wykrywają one sygnatury cieplne obiektów, umożliwiając systemom AI wykrywanie i śledzenie ruchu w ciemności.



Rysunek 1: Czujnik podczerwieni

Czujniki dotyku: Czujniki dotyku są wykorzystywane w aplikacjach takich jak ekrany dotykowe, chwytaki robotów i protezy. Czujniki dotykowe mogą wykrywać ciśnienie, temperaturę i inne właściwości fizyczne, umożliwiając systemom sztucznej inteligencji interakcję ze światem fizycznym.



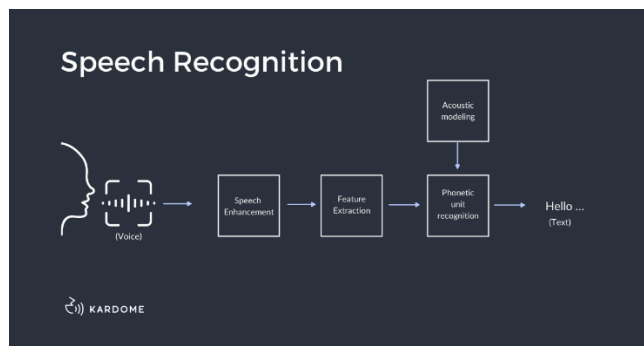
Rysunek 2: Czujnik dotykowy

4.3 Pomysł 2: Reprezentacja i rozumowanie

Reprezentacja odnosi się do sposobu przechowywania i przetwarzania danych przez systemy sztucznej inteligencji. Reprezentacja ta pozwala systemowi sztucznej inteligencji analizować i wyodrębniać cechy dźwięku, takie jak fonemy lub słowa. W sztucznej inteligencji dane są reprezentowane w sposób, który maszyny mogą zrozumieć i skutecznie przetwarzać. Najczęstszym sposobem reprezentacji danych w sztucznej inteligencji są symbole lub liczby. Na przykład zdjęcie kota może być reprezentowane jako macierz liczb, które odpowiadają kolorowi każdego piksela na obrazie. Rozumowanie to proces wyciągania wniosków z zebranych i przedstawionych danych. Systemy sztucznej inteligencji wykorzystują różne techniki rozumowania do przetwarzania i analizowania danych, takie jak rozumowanie dedukcyjne, rozumowanie indukcyjne i rozumowanie abdukcyjne. Techniki te pomagają maszynom przewidywać, rozpoznawać wzorce i rozwiązywać problemy.

Przykłady:

Rozpoznawanie mowy: Systemy AI reprezentują język mówiony jako sekwencję symboli lub liczb, a następnie wykorzystują techniki rozumowania, aby zinterpretować i zrozumieć znaczenie mowy. System sztucznej inteligencji może wykorzystywać rozumowanie w celu dopasowania danych audio do bazy danych znanych słów lub fraz lub w celu zidentyfikowania intencji wypowiedzanego polecenia.



Rysunek 3: Rozpoznawanie mowy

Wizja komputerowa: Systemy sztucznej inteligencji reprezentują obrazy jako macierze liczb, a następnie wykorzystują techniki rozumowania do rozpoznawania obiektów i wzorców na obrazach. Rozumowanie może obejmować wykorzystanie algorytmów rozpoznawania obrazu do identyfikacji obiektów w scenie lub do śledzenia ruchu obiektów w czasie.

Dziedzina medycyny: Sztuczna inteligencja może analizować duże ilości danych pacjentów w celu identyfikacji wzorców i przewidywania potencjalnych problemów zdrowotnych. Może to pomóc lekarzom w podejmowaniu bardziej świadomych decyzji dotyczących opieki nad pacjentem.

Finanse: Sztuczna inteligencja może analizować trendy rynkowe i dane w celu identyfikacji wzorców i przewidywania przyszłych ruchów na rynku. Może to pomóc inwestorom w podejmowaniu bardziej świadomych decyzji dotyczących inwestycji i zarządzaniu ryzykiem.

4.4 Pomysł 3: Uczenie się

Uczenie się jest kluczowym aspektem sztucznej inteligencji i odróżnia ją od tradycyjnego programowania komputerowego. Algorytmy uczenia maszynowego umożliwiają systemom AI uczenie się i doskonalenie na podstawie doświadczenia bez konieczności ich wyraźnego programowania. W kontekście aplikacji mowy i wizji uczenie się odgrywa kluczową rolę w umożliwianiu maszynom rozpoznawania wzorców i przewidywania na podstawie danych. Znaczenie uczenia się w aplikacjach mowy i wizji jest nie do przecenienia. Ucząc się na podstawie dużych ilości danych, maszyny mogą rozpoznawać złożone wzorce i dokonywać dokładniejszych prognoz, co prowadzi do poprawy wydajności w rzeczywistych zastosowaniach, takich jak rozpoznawanie mowy, wykrywanie obiektów i autonomiczna jazda.

Trzy główne typy uczenia maszynowego obejmują:

Uczenie nadzorowane: Uczenie nadzorowane jest wykorzystywane do zadań takich jak rozpoznawanie mowy i obrazów. Uczenie nadzorowane to rodzaj uczenia maszynowego, w którym algorytm jest trenowany na etykietowanym zbiorze danych. Oznakowany zbiór danych składa się z danych wejściowych (np. obrazów lub plików audio) wraz z odpowiednimi etykietami wyjściowymi (np. klasyfikacja tekstu lub obrazu). Celem algorytmu jest nauczenie się funkcji mapowania z danych wejściowych na etykiety wyjściowe. W rozpoznawaniu mowy uczenie nadzorowane jest wykorzystywane do trenowania algorytmu w celu rozpoznawania wypowiedzianych słów i konwertowania ich na tekst. Dane wejściowe składają się z plików audio, a etykiety wyjściowe to odpowiadające im transkrypcje. Algorytm jest trenowany na dużym zbiorze danych etykietowanych plików audio i transkrypcji i uczy się rozpoznawać wzorce i relacje między danymi audio a odpowiadającym im tekstem. Niektóre popularne algorytmy uczenia nadzorowanego stosowane w aplikacjach mowy i wizji to konwolucyjne sieci neuronowe (CNN) i rekurencyjne sieci neuronowe (RNN). CNN są szczególnie dobrze przystosowane do zadań rozpoznawania obrazu, ponieważ mogą automatycznie uczyć się hierarchicznych reprezentacji cech wizualnych. Sieci RNN szczególnie dobrze nadają się do zadań rozpoznawania mowy, ponieważ mogą modelować zależności czasowe między danymi audio a odpowiadającym im tekstem.

Uczenie bez nadzoru: Uczenie bez nadzoru to rodzaj uczenia maszynowego, który obejmuje uczenie się wzorców i relacji w danych bez potrzeby stosowania jawnych etykiet lub nadzoru. Uczenie bez nadzoru jest szczególnie przydatne w zastosowaniach wizji komputerowej, gdzie duże ilości nieustrukturyzowanych danych wizualnych muszą być analizowane w celu zidentyfikowania znaczących wzorców i relacji. Na przykład algorytmy grupowania mogą być używane do grupowania podobnych obrazów lub obiektów w zbiorze danych w oparciu o ich cechy wizualne, bez potrzeby jawnego etykietowania. Innym przykładem algorytmu uczenia się bez nadzoru jest autoenkoder. Autoenkoder to sieć neuronowa, która jest trenowana w celu rekonstrukcji danych wejściowych z niższej wymiarowej reprezentacji tych danych. Część kodująca sieci redukuje dane wejściowe do przestrzeni ukrytej o niższym wymiarze, a część dekodująca sieci próbuje zrekonstruować oryginalne dane wejściowe z tej ukrytej reprezentacji. Autoenkodery mogą być wykorzystywane do zadań takich jak odszumianie obrazu, redukcja wymiarowości i uczenie się cech. Innym przykładem algorytmu uczenia się bez nadzoru jest generatywna sieć przeciwstawna (GAN).

Uczenie ze wzmocnieniem: W zastosowaniach związanych z mową i widzeniem, uczenie ze wzmocnieniem jest wykorzystywane w takich obszarach jak robotyka, autonomiczna jazda i gry. Na przykład, uczenie ze wzmocnieniem może być wykorzystywane do uczenia robota poruszania się w zagrożonym środowisku lub do kontrolowania ruchów samojezdnego samochodu. W grach uczenie ze wzmocnieniem może być wykorzystywane do opracowywania agentów, którzy mogą grać w gry takie jak szachy lub Go na nadludzkim poziomie. Jednym z najpopularniejszych algorytmów uczenia ze wzmocnieniem jest uczenie Q, które wykorzystuje tabelę Q do przechowywania oczekiwanej nagrody za każde działanie w każdym stanie. Agent aktualizuje tabelę Q w oparciu o otrzymane nagrody i wykorzystuje zaktualizowaną tabelę do wyboru działań w przyszłości.

Actions : ↑ → ↓ ←

Start	0	0	0	0
Nothing / Blank	0	0	0	0
Power	0	0	0	0
Mines	0	0	0	0
END	0	0	0	0

Rysunek 4: Tabela Q

4.5 Pomysł 4: Naturalna interakcja

Naturalna interakcja w sztucznej inteligencji i robotach to zdolność do rozumienia i reagowania na ludzki język naturalny i działania, takie jak głos i gesty. Jest to szczególnie istotne w zastosowaniach mowy i wizji, gdzie celem jest stworzenie interfejsów, które mogą zrozumieć ludzki język i wskazówki wizualne oraz reagować w sposób zrozumiały, intuicyjny i użyteczny. Technologie interakcji mowy i wizji, takie jak rozpoznawanie mowy, rozpoznawanie gestów i rozpoznawanie twarzy, zmieniają sposób interakcji z maszynami i otwierają nowe możliwości zastosowań w grach, robotyce, bezpieczeństwie i nie tylko.

Interakcja mowy: Interakcja mowy polega na wykorzystaniu języka naturalnego do komunikacji między ludźmi a systemami sztucznej inteligencji i obejmuje rozpoznawanie i syntezę mowy. Rozpoznawanie mowy przekształca język mówiony w tekst za pomocą algorytmów, podczas gdy synteza mowy generuje sztuczną mowę z tekstu za pomocą systemów TTS. Jednym z głównych zastosowań interakcji mowy w sztucznej inteligencji są wirtualni asystenci, tacy jak Siri firmy Apple, Alexa firmy Amazon i Asystent Google. Ci wirtualni asystenci wykorzystują rozpoznawanie mowy i NLP, aby zrozumieć żądania użytkowników i odpowiedzieć odpowiednimi informacjami lub działaniami. Interakcja mowy jest również wykorzystywana w centrach telefonicznych, gdzie zautomatyzowane systemy głosowe mogą kierować połączenia i odpowiadać na podstawowe pytania, uwalniając ludzkich agentów do bardziej złożonych zadań.

Interakcja wzrokowa: Interakcja wzrokowa odnosi się do sposobów, w jakie ludzie mogą wchodzić w interakcje z systemami sztucznej inteligencji za pomocą wskazówek wizualnych, gestów i innych niewerbalnych metod komunikacji. Obejmuje to technologie takie jak rozpoznawanie twarzy, śledzenie wzroku i przechwytywanie ruchu. Technologie te mają wiele praktycznych zastosowań, takich jak sterowanie urządzeniami za pomocą gestów dłoni, śledzenie ruchów gałek ocznych oraz rozpoznawanie wyrazu twarzy i emocji.

Wirtualna i rozszerzona rzeczywistość: Wirtualna i rozszerzona rzeczywistość (VR i AR) to technologie wykorzystywane w sztucznej inteligencji do tworzenia wciągających interakcji między ludźmi i maszynami. VR tworzy w pełni wirtualne środowisko przy użyciu zestawów słuchawkowych i kontrolerów, podczas gdy AR nakłada wirtualne obiekty na widok rzeczywistego świata użytkownika. Technologie te mają swoje ograniczenia, takie jak konieczność posiadania specjalistycznego sprzętu i możliwość wystąpienia choroby lokomocyjnej, ale ewoluują i mają liczne zastosowania w dziedzinach takich jak edukacja, opieka zdrowotna i rozrywka.



Rysunek 5: Wirtualna rzeczywistość



Rysunek 6: Rozszerzona rzeczywistość

4.6 Pomysł 5: Wpływ społeczny

Spółeczny wpływ sztucznej inteligencji w zastosowaniach mowy i wizji stał się w ostatnich latach kluczowym tematem dyskusji. Chociaż technologie te mogą potencjalnie przynieść znaczący postęp w różnych dziedzinach, mają one również konsekwencje etyczne, społeczne i ekonomiczne, którymi należy się zająć.

Zalety sztucznej inteligencji w zastosowaniach mowy i wizji: Sztuczna inteligencja w mowie i wizji ma wady i zalety dla społeczeństwa. Pozytywnym aspektem jest poprawa opieki zdrowotnej, edukacji i bezpieczeństwa publicznego. Na przykład rozpoznawanie mowy pomaga lekarzom w dokładnej transkrypcji dokumentacji medycznej, a technologie wspomagające pomagają osobom niepełnosprawnym. Aplikacje wizyjne przyczyniają się do rozpoznawania twarzy i autonomicznych pojazdów. Odpowiedzialny rozwój i wdrażanie technologii AI wymaga jednak uwzględnienia kwestii etycznych.

Wady sztucznej inteligencji w zastosowaniach mowy i wizji: Sztuczna inteligencja w mowie i wizji ma potencjalne wady. Jedną z głównych obaw jest potencjał uprzedzeń i dyskryminacji, ponieważ algorytmy AI mogą powielać uprzedzenia społeczne. Istnieje również ryzyko przenoszenia miejsc pracy i nierówności ekonomicznych ze względu na wysokie koszty systemów AI. Prywatność i nadzór to kwestie etyczne, ponieważ sztuczna inteligencja gromadzi dane osobowe. Ponadto istnieją szersze obawy dotyczące długoterminowego wpływu sztucznej inteligencji, ponieważ staje się ona coraz bardziej zaawansowana i autonomiczna. Konieczne jest zajęcie się tymi obawami i stworzenie bardziej przejrzystych, sprawiedliwych i odpowiedzialnych systemów, które przyniosą korzyści społeczeństwu.

Kwestie etyczne: Podczas gdy technologie AI mogą potencjalnie przynieść znaczące korzyści, istnieje również kilka kwestii etycznych, które należy wziąć pod uwagę. Niektóre z kluczowych kwestii etycznych związanych ze sztuczną inteligencją w zastosowaniach mowy i wizji obejmują prywatność, przejrzystość, stronniczość i odpowiedzialność.

- **Prywatność:** Wykorzystanie sztucznej inteligencji w aplikacjach mowy i wizji budzi obawy dotyczące prywatności. Wraz z rosnącą ilością gromadzonych danych istnieje ryzyko, że dane osobowe mogą zostać niewłaściwie wykorzystane lub niewłaściwie potraktowane. W związku z tym ważne jest, aby programiści i organizacje zapewniły odpowiednią ochronę prywatności w celu ochrony danych osobowych.
- **Przejrzystość:** Kolejną kwestią etyczną związaną ze sztuczną inteligencją w aplikacjach mowy i wizji jest kwestia przejrzystości. Zrozumienie, w jaki sposób system sztucznej inteligencji

podejmuje decyzje, może być trudne, co może utrudniać ocenę uczciwości i dokładności jego wyników. Aby rozwiązać tę kwestię, ważne jest, aby programiści priorytetowo traktowali przejrzystość swoich systemów sztucznej inteligencji i upewnili się, że można je skontrolować i wyjaśnić.

- **Stronniczość:** Stronniczość to kolejna kwestia etyczna, która może pojawić się w zastosowaniach AI w mowie i wizji. Jeśli dane wykorzystywane do szkolenia systemu sztucznej inteligencji są stronnicze, wyniki systemu również mogą być stronnicze. Może to mieć znaczące konsekwencje w obszarach takich jak zatrudnianie, udzielanie pożyczek i wymiar sprawiedliwości w sprawach karnych. Aby rozwiązać tę kwestię, programiści muszą podjąć kroki w celu zapewnienia, że ich zestawy danych są reprezentatywne i nieobiektywne.
- **Odpowiedzialność:** Wreszcie, wykorzystanie sztucznej inteligencji w aplikacjach mowy i wizji rodzi pytania o odpowiedzialność. Kto ponosi odpowiedzialność, jeśli system AI popełni błąd lub wyrządzi szkodę? W miarę jak sztuczna inteligencja staje się coraz bardziej zintegrowana z naszym życiem, ważne jest, aby upewnić się, że istnieją jasne linie odpowiedzialności i odpowiedzialności.

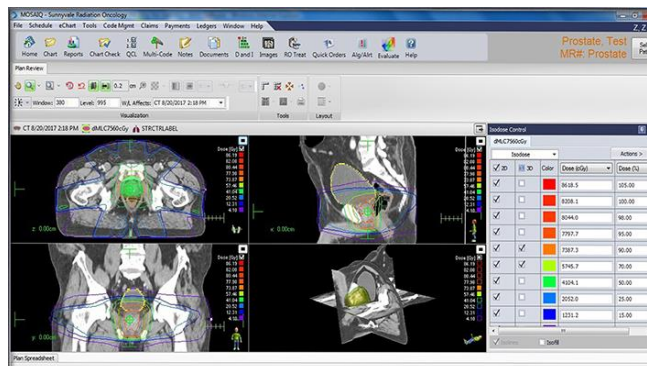
4.7 Studia przypadków i historie sukcesu

Watson for Oncology firmy IBM pomaga onkologom w tworzeniu spersonalizowanych planów leczenia raka. Amazon's Alexa for Speech Recognition zrewolucjonizowała sposób, w jaki ludzie wchodzi w interakcje z technologią. Inne godne uwagi zastosowania sztucznej inteligencji w mowie i wizji obejmują Face ID firmy Apple i Autopilota Tesli. Przykłady te pokazują transformacyjny potencjał sztucznej inteligencji w ulepszaniu naszego życia.

4.7.1 Przykład 1: Sztuczna inteligencja Google DeepMind do czytania z ruchu warg

DeepMind firmy Google stworzył system sztucznej inteligencji przeszkolony na 5000 godzin materiału telewizyjnego, który osiągnął dokładność 46,8% w czytaniu z ruchu warg, przewyższając 12,4% ludzkich ekspertów. Ma to potencjalne zastosowanie w rozpoznawaniu mowy, bezpieczeństwie i pomocy osobom z upośledzeniem słuchu. Czytanie z ruchu warg jest trudne, a nawet najlepsi ludzie potrafią dokładnie zrozumieć tylko około 30% wypowiedzanych słów. Sztuczna inteligencja DeepMind może zrewolucjonizować komunikację dla osób niedosłyszących. Istnieją obawy dotyczące prywatności związane z gromadzeniem i analizą dużych ilości nagrań wideo oraz możliwością niewłaściwego wykorzystania w kontekście nadzoru i egzekwowania prawa. Ogólnie rzecz biorąc, technologia ta pokazuje, w jaki sposób sztuczna inteligencja może poprawić komunikację dla osób z upośledzeniem słuchu i podkreśla potrzebę rozważań etycznych w zakresie rozwoju i wykorzystania sztucznej inteligencji.

4.7.2 Przykład 2: IBM Watson dla onkologii



Rysunek 7: IBM Watson

IBM Watson for Oncology to system sztucznej inteligencji, który pomaga onkologom w podejmowaniu spersonalizowanych decyzji dotyczących leczenia pacjentów chorych na raka. System wykorzystuje przetwarzanie języka naturalnego i uczenie maszynowe do analizowania ogromnych ilości danych medycznych i dostarczania zaleceń dotyczących leczenia w oparciu o unikalne okoliczności pacjenta. Mimo pewnych wyzwań i krytyki systemu, był on z powodzeniem stosowany w placówkach opieki zdrowotnej i stanowi obiecujący przykład tego, jak sztuczna inteligencja może poprawić wyniki leczenia pacjentów.

4.7.3 Przykład 3: Alexa firmy Amazon do rozpoznawania mowy

Alexa firmy Amazon to popularny wirtualny asystent, który wykorzystuje technologię rozpoznawania mowy, aby umożliwić obsługę różnych urządzeń bez użycia rąk. Rozpoznaje i reaguje na polecenia w języku naturalnym i wykorzystuje uczenie maszynowe do dostarczania spersonalizowanych rekomendacji. Alexa pomogła również osobom niepełnosprawnym, umożliwiając im łatwiejsze wykonywanie codziennych zadań. Należy jednak zająć się kwestiami prywatności i bezpieczeństwa. Sukces Alexy pokazuje potencjał technologii rozpoznawania mowy w zakresie poprawy naszego codziennego życia, ale należy wziąć pod uwagę kwestie etyczne i ryzyko związane ze sztuczną inteligencją.

5 Dodatkowe materiały i zasoby

Rodzaj zasobu	Tytuł	Temat	Link
Artykuł	"Rozpoznawanie mowy: Przyszłość jest teraz"	Rozpoznawanie mowy	https://www.wired.com/story/speech-recognition-future-is-now/

Artykuł	"Compute Vision: Przyszłość sztucznej inteligencji"	Wizja komputerowa	https://www.forbes.com/sites/forbestechcouncil/2021/01/05/computer-vision-the-future-of-artificial-intelligence/?sh=39c773b742f4
Artykuł	"Jak głębokie uczenie zmienia przyszłość treści wizualnych"	Uczenie głębokie	https://blog.hubspot.com/marketing/deep-learning-visual-content
Artykuł	"Plusy minusy sztucznej inteligencji w opiece zdrowotnej"	Sztuczna inteligencja w opiece zdrowotnej	https://www.healthtechmagazine.net/article/2020/03/pros-and-cons-ai-healthcare
Wideo	"TED Talk: Jak sztuczna inteligencja ułatwia diagnozowanie chorób"	Zalety wizji i mowy w opiece zdrowotnej	https://www.ted.com/talks/pratik_shah_how_ai_is_making_it_easier_to_diagnose_disease
Wideo	"Jak sztuczna inteligencja zmienia nasz sposób postrzegania świata"	Zmieniające świat efekty rozwoju i wykorzystania sztucznej inteligencji	https://www.youtube.com/watch?v=BDBTJOGvCv4
Link	TensorFlow	TensorFlow: platforma uczenia maszynowego o otwartym kodzie źródłowym do tworzenia modeli uczenia głębokiego	https://www.tensorflow.org/

Link	OpenCV	OpenCV: biblioteka wizji komputerowej typu open source do tworzenia aplikacji wizji komputerowej w czasie rzeczywistym	https://opencv.org/
Link	PyTorch	PyTorch: biblioteka uczenia maszynowego typu open source do tworzenia modeli głębokiego uczenia się modeli	https://pytorch.org/
Link	NVIDIA	NVIDIA Deep Learning Institute	https://www.nvidia.com/en-us/deep-learning-ai/education/

6 Podsumowanie

Podsumowując, jednostka dotycząca zastosowania sztucznej inteligencji w mowie i wizji obejmowała różne tematy, w tym przetwarzanie języka naturalnego, wizję komputerową, uczenie maszynowe i wpływ społeczny. Przeanalizowaliśmy studia przypadków, takie jak DeepMind AI firmy Google do czytania z ruchu warg, Watson firmy IBM do onkologii i Alexa firmy Amazon do rozpoznawania mowy, które ilustrują potencjał sztucznej inteligencji w rozwiązywaniu rzeczywistych problemów. Należy jednak dokładnie przeanalizować kwestie etyczne oraz potencjał stronniczości i ograniczeń sztucznej inteligencji. Dzięki tej wiedzy uczniowie mogą nadal odkrywać ogromne możliwości i wyzwania w dziedzinie sztucznej inteligencji.

7 Quiz

Pytanie 1: Prawda/Fałsz

- Uczenie głębokie jest podzbiorem uczenia maszynowego. (**Prawda/Fałsz**)
- Stronniczość i nadmierne dopasowanie to typowe wyzwania w uczeniu maszynowym. (**Prawda/Fałsz**)

Pytanie 2: Dopasuj następujące typy uczenia maszynowego do ich opisów

a. Uczenie nadzorowane	i. Uczenie się na podstawie oznaczonych danych
b. Uczenie się bez nadzoru	ii. Uczenie się bez oznaczonych danych
c. Uczenie ze wzmocnieniem	iii. Uczenie się z systemem nagród

Odpowiedź: a.i, b.ii, c.iii

Pytanie 3: Dopasuj następujące algorytmy uczenia maszynowego do ich zastosowań:

a. Konwolucyjne sieci neuronowe (KSN)	i. Rozpoznawanie mowy
b. Autoenkodery	ii. Klasyfikacja obrazów
c. Uczenie Q	iii. Klasyfikacja obrazów

Odpowiedź: a.ii, b.i, c.iii

Pytanie 4: Jaka jest główna różnica między uczeniem nadzorowanym a nienadzorowanym?

- a) **Uczenie nadzorowane wymaga oznaczonych danych, podczas gdy uczenie nienadzorowane nie.**
- b) Uczenie bez nadzoru wymaga oznaczonych danych, podczas gdy uczenie nadzorowane nie.
- c) Oba rodzaje uczenia się wymagają oznaczonych danych.
- d) Nie ma różnicy między uczeniem nadzorowanym i nienadzorowanym.

Pytanie 5: Który rodzaj uczenia się jest wykorzystywany w robotyce i grach?

- a) Uczenie nadzorowane
- b) Uczenie bez nadzoru
- c) **Uczenie ze wzmocnieniem**
- d) Żadne z powyższych

Pytanie 6: Jakie jest główne ograniczenie uczenia maszynowego w aplikacjach mowy i wizji?

- a) Brak mocy obliczeniowej
- b) Brak oznaczonych danych
- c) **Stronniczość i nadmierne dopasowanie**
- d) Brak możliwości interpretacji

Pytanie 7: Jaka jest główna zaleta korzystania z Deep Learning?

- a) **Lepsza dokładność w złożonych zadaniach**
- b) Niższe wymagania obliczeniowe
- c) Mniejsza podatność na nadmierne dopasowanie
- d) Łatwiejsze do interpretacji

Pytanie 8: Jaki jest główny czynnik etyczny w rozwoju sztucznej inteligencji pod kątem wpływu na społeczeństwo?

- a) Zapewnienie bezpieczeństwa i ochrony użytkowników
- b) Zapewnienie sprawiedliwości i unikanie stronniczości**
- c) Zapewnienie rentowności dla firmy
- d) Zapewnienie kompatybilności z istniejącymi systemami

Pytanie 9: Jaka jest główna zaleta korzystania z rozszerzonej rzeczywistości w połączeniu ze sztuczną inteligencją?

- a) Poprawa doświadczenia użytkownika**
- b) Zmniejszenie wymagań obliczeniowych
- c) Poprawa dokładności złożonych zadań
- d) Zwiększenie możliwości interpretacji sztucznej inteligencji

8 Referencje

- Kelleher, J. D., & Tierney, B. (2018). Data science: An introduction. CRC Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
- Jurafsky, D., & Martin, J. H. (2020). Speech and language processing (3rd ed.). Pearson.
- Zhang, L., Tan, T., & Li, Y. (2016). Deep learning-based face recognition: A survey. International Journal of Automation and Computing, 13(4), 261-279.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484- 489.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533.
- Weng, L., Zhang, Y., Xue, X., & Chen, K. (2021). The applications of artificial intelligence in computer vision. In 2021 International Conference on Computer Science and Artificial Intelligence (CSAI) (pp. 139-143). IEEE.
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).
- Dhillon, G., & Moores, D. (2018). Natural language processing in AI: past, present and future. In Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval (pp. 1-10).
- Canny, J. (2021). Ethics and artificial intelligence. Communications of the ACM, 64(1), 18-20.
- Li, X., Liang, J., & Li, Y. (2021). Research on the impact of augmented reality technology on the experience of cultural tourism. Journal of Physics: Conference Series, 1838(1), 012064.
- Yaqoob, I., Ahmed, M. M., Gani, A., Imran, M., Guizani, M., & Hitha, H. (2019). Virtual reality for education: A survey. Journal of Network and Com