

Teachable Machine – Trenuj i poprawiaj model AI

Warsztat ma na celu zapoznanie uczestników z narzędziem Teachable Machine oraz z podstawowymi zasadami uczenia maszynowego. Uczestnicy samodzielnie tworzą model rozpoznający obrazy, testują go w praktyce, a następnie – co stanowi sedno warsztatu – świadomie poprawiają, identyfikując typowe błędy w danych treningowych. Celem warsztatu jest nie tylko poznanie sposobów działania i trenowania sztucznej inteligencji, ale również uwrażliwienie uczestników na błędy, jakie AI popełnia, gdy dane treningowe są niewystarczające lub stronnicze. Zdobyta wiedza o jakości danych jest jednocześnie bezpośrednią podstawą kryteriów oceny modeli w Etapie I Konkursu ścieżki AI.

Czas i odbiorcy:

Czas trwania: 90 min.

Grupa wiekowa: od 14 do 19 lat

Cele:

Cele ogólne:

- Poznanie zasad działania uczenia maszynowego i metod trenowania AI.
- Kształtowanie umiejętności posługiwania się narzędziem Teachable Machine.
- Budowanie świadomości błędów towarzyszących trenowaniu i działaniu sztucznej inteligencji.
- Przygotowanie uczestników do zadania konkursowego w ścieżce AI.

Cele szczegółowe:

- Samodzielne stworzenie modelu klasyfikacji obrazów w Teachable Machine.
- Zidentyfikowanie co najmniej kilku typowych błędów w danych treningowych.
- Zastosowanie konkretnych poprawek i zaobserwowanie ich wpływu na jakość modelu.
- Omówienie głośnych wpadek AI z mediów i powiązanie ich z poznanymi błędami.
- Zrozumienie, w jaki sposób błędy w danych treningowych przekładają się na realne konsekwencje.

Potrzebne materiały:

Komputery z kamerą i dostępem do Internetu, kartki papieru A4, długopisy.

Przygotowanie dla trenera:

Zapoznaj się z narzędziem Teachable Machine, wchodząc na stronę teachablemachine.withgoogle.com. Przejdź samodzielnie cały proces: wybierz Image Project, zbierz kilkadziesiąt zdjęć dwóch klas (np. Rozpoznawanie gestów papier, kamień, norzyce.), wytrenuj model i przetestuj go na żywo. Upewnij się, że kamery w komputerach uczestników działają przed warsztatem – Teachable Machine korzysta z kamery i bez niej ćwiczenie nie jest utrudnione (należałoby wcześniej przygotować pakiety zdjęć, a następnie je zaimportować). Przygotuj dwa wyraźne rekwizyty widoczne na tle sali – np. pomarańcza i banan, lub dowolne dwa wyraźnie różne przedmioty. Warto mieć dostęp do co najmniej dwóch różnych tła: np. biała kartka i ciemna kurtka położona na biurku – przyda się do demonstracji błędu nr 3. Zapoznaj się z opisanymi poniżej case'ami z mediów i przygotuj je do wyświetlenia lub wydruku. Przygotuj też dołączoną prezentację błędów w danych treningowych i sposobów ich poprawy – posłużą one do szybkiego przeglądu błędów w bloku „Teoria błędów”, a uczestnicy będą mogli z nich korzystać także później, np. podczas pracy nad zadaniem konkursowym (nauczyciel może im taką prezentację przestać po warsztacie).

ETAPY REALIZACJI:

Czas	Opis ćwiczenia	Potrzebne materiały / uwagi
10 min	<u>Wprowadzenie – czym jest uczenie maszynowe?</u> Przywitaj uczestników na warsztatach. Zapytaj, czy byli już na jakichś warsztatach w CDT (jeśli tak, to jakich) i czy wiedzą, na jakie warsztaty przyszli dzisiaj. Powiedz, że dzisiaj sami wytrenujemy modele i – co najważniejsze, będziemy je świadomie poprawiać. Wytłumacz, że te warsztaty są elementem nowej ścieżki AI, zakończonej ogólnopolskim konkursem, którego celem jest stworzenie prototypu robota ze sztuczną inteligencją.	teachablemachine.withgoogle.com Komputer z rzutnikiem lub tablica multimedialna

	Ustal z uczestnikami zasady towarzyszące nam podczas dzisiejszego warsztatu, między innymi: - zasada 3+10 (<i>Czasem poproszę Was o 3 minuty pełnego skupienia. Jest to dla mnie ważne, ponieważ przez ten czas będę tłumaczyć Wam kolejne zadania lub sposób wykonania kolejnego ćwiczenia. Jeśli przez te 3 minuty nie będę mieć Waszej uwagi, to potem nie będziecie wiedzieli co i jak robić. Jest to ważne, bo po trzech minutach dla mnie jest 10, 15, a czasem nawet 20 minut dla Was na samodzielną pracę. Muszę mieć zatem Waszą uwagę przez 3 minuty, żebyście potem mogli bez problemu pracować przez 20.</i>) - zasada pauzy (zatrzymywanie pracy i przemykanie kłap komputerów) - zasada dobrej współpracy (nie krytykujemy, nie wyśmiewamy, pomagamy sobie wzajemnie i dzielimy się pracą) Następnie przeprowadź krótkie wprowadzenie do tematu. Zadaj grupie pytania i wysłuchaj odpowiedzi: - Skąd Spotify wie, jakiej muzyki lubisz słuchać? - Skąd kamera w telefonie wie, że na zdjęciu jest twarz? - Skąd ChatGPT umie pisać po polsku? - Skąd system w sklepie wie, że interesują Cię właśnie te buty, a nie tamte? Docień odpowiedzi bliskie tematowi i naprowadź na pojęcie danych treningowych. Powiedz: <i>Wszystkie te systemy działają dlatego, że ktoś pokazał im ogromną liczbę przykładów. Spotify analizowało miliony piosenek słuchanych przez miliony ludzi. Rozpoznawanie twarzy było trenowane na setkach milionów</i>	
--	---	--

	zdjęć. ChatGPT przetworzył prawie cały Internet. AI nie myśli, szuka wzorców w danych. Im lepsze dane, tym lepszy model. I odwrotnie, im gorsze dane, tym większy problem. Dzisiaj będziecie to mogli zobaczyć na własne oczy: wytrenujecie swój model, przetestujecie go i będziecie go poprawiać. Pokaż na rzutniku ekran strony Teachable Machine. Wyjaśnij krótko trzy dostępne tryby: Image (obraz z kamery), audio (dźwięk z mikrofonu), pozycja (układ ciała, wzajemna pozycja punktów w przestrzeni). Zaznacz, że dziś pracujemy wyłącznie z trybem Image. Powiedz, że Teachable Machine to narzędzie Google, które pozwala wytrenować własny model klasyfikacji obrazów bezpośrednio w przeglądarce, bez pisania kodu. Wyszkolony model można potem wyeksportować i używać np. do sterowania robotem lub w aplikacji internetowej.	
15 min	Pierwsze podejście – tworzenie własnego modelu Poproś uczestników o otwarcie strony Teachable Machine w przeglądarce. Pokazuj kolejne kroki na rzutniku, a uczestnicy wykonują je równocześnie na swoich komputerach: - Krok 1: kliknij <i>Get Started</i>, następnie <i>Image Project</i>, następnie <i>Standard image model</i>. - Krok 2: domyślnie są dwie klasy: Class 1 i Class 2. Zmień ich nazwy na nazwy wybranych obiektów, np. Banan i Pomarańcza, albo długopis i linijka. Nazwę zmienia się klikając ołówek obok nazwy klasy. - Krok 3: kliknij <i>Webcam</i> przy pierwszej klasie. Przytrzymaj przycisk <i>Hold to Record</i> (przytrzymaj, aby nagrać) – TM automatycznie robi serię zdjęć co kilka klatek. Zbierajcie	Komputery z kamerą Rekwizyty uczniów (owoce, okulary itp.) <i>Jeśli kamera nie działa:</i> sprawdź uprawnienia w ustawieniach przeglądarki (ikona kamery w pasku adresu). teachablemachine.withgoogle.com Uwaga: do korzystania z teachablemachine potrzebny jest komputer z kamerką. Strona ta nie działa na urządzeniach mobilnych. Na części urządzeń mobilnych możliwe jest uruchomienie starej wersji teachable machine: https://teachablemachine.withgoogle.com/v1 jednak może być ona wolna i niestabilna.

	zdjęcia obiektu, poruszając nim powoli przed kamerą. Cel: około 50–60 zdjęć na klasę. Powtórz dla wszystkich klas. - Krok 4: po zebraniu zdjęć dla wszystkich klas kliknij <i>Train Model</i> i poczekaj na zakończenie treningu (pasek postępu w panelu środkowym). - Krok 5: panel <i>Preview (podgląd)</i> po prawej stronie pokazuje na żywo, jakie klasy rozpoznaje kamera, z procentowym wskaźnikiem pewności. Przetestujcie model, pokazując przed kamerą różne obiekty. Każda para lub trójka wybiera jeden z poniższych zestawów do rozróżniania. Ważne: obiekty muszą być fizycznie dostępne przy komputerze. - długopis vs linijka - pomarańcza vs banan (lub inny owoc dostępny w sali) - twarz z okularami vs twarz bez okularów - własny pomysł – musi być fizycznie dostępny przy komputerze Daj grupom 12–13 minut na zebranie zdjęć i trening. Chodząc po sali, sprawdzaj czy wszystkie grupy zebrały wystarczającą liczbę zdjęć i czy trening zakończył się poprawnie. Zwróć uwagę na grupy, które zebrały mniej niż 30 zdjęć na klasę – poproś je o uzupełnienie. Ważne: <i>nie zamykać karty przeglądarki podczas treningu – model trenuje lokalnie w przeglądarce i utrata strony oznacza konieczność zaczęcia od nowa.</i>	
--	---	--

	Po zakończeniu treningu przez wszystkie grupy nie organizuj jeszcze swobodnego testowania – przejdźcie od razu do kolejnego, ustrukturyzowanego ćwiczenia „ Testerzy AI ”.	
10 min	Testerzy AI Poproś każdą grupę, aby w jednym zdaniu powiedziała, co rozpoznaje jej model (np. „Nasz model rozpoznaje banana i pomarańczę”). Wszyscy uważnie słuchają i zapamiętują, bo będzie to miało znaczenie w kolejnym zadaniu. Następnie zdecyduj, w jaki sposób grupy się wymieniają, dobierz je w pary albo poproś, aby wszystkie grupy jednocześnie „rotowały” zgodnie z ruchem wskazówek zegara do stanowiska (komputera) sąsiedniej grupy. Na nowym stanowisku każda grupa testuje model innej grupy, pokazuje mu obiekty, sprawdza różne ułożenia, odległości i tła, i zapisuje na kartce listę zaobserwowanych „błędów”, czyli sytuacji, w których model się mylił. Podkreśl, aby grupy pisały czytelnie. Grupy zostają przy testowanym modelu – nie wracają jeszcze do swoich komputerów, bo lista błędów posłuży za chwilę do wspólnej dyskusji.	Kartki papieru A4, długopisy
10 min	Test i wspólna obserwacja błędów Grupy zostają na stanowisku, przy którym właśnie testowały model innej grupy, nie wracają jeszcze do swoich komputerów. Zatrzymaj grupę i przeprowadź krótką dyskusję na podstawie tego, co zaobserwowali podczas testowania (odnośnie modelu innej grupy). Zadaj pytania i zbieraj odpowiedzi na tablicy: - Czy testowany przez Was model się mylił? Co wtedy było inne niż w bazie zdjęć?	Tablica lub kartka do zapisywania obserwacji grupy

	- Czy model działał tak samo przy testowaniu przez inne osoby? - Czy model działał, gdy schowaliście obiekt i pokazaliście samo tło? - Co byście zmienili w zebranych danych, gdybyście mieli coś rekomendować drugiej drużynie? Zbieraj odpowiedzi na tablicy lub kartce. Szukaj wśród nich sformułowań związanych z: - tłem na zdjęciach (inny kolor ściany, inne miejsce) - oświetleniem lub porą dnia - różnymi osobami lub różnymi rękami - kątem fotografowania lub odległością od kamery - modelem, który „widzi” obiekt nawet gdy go nie ma Nie oceniaj jeszcze ani nie podpowiadaj nazw błędów – chodzi o to, żeby uczestnicy sami sformułowali problemy. Gdy lista obserwacji jest gotowa, powiedz: <i>Świetnie. Te obserwacje to właśnie podstawa tego, o czym będziemy rozmawiać w następnej części lekcji. Zobaczycie zaraz, że te problemy mają nazwy i że mają je wszystkie systemy AI na świecie, nie tylko wasze modele. Dokładnie te same błędy pojawiają się w systemach rozpoznawania twarzy używanych przez policję, w diagnostyce medycznej i w samochodach autonomicznych.</i> Dopiero teraz każda grupa pozostawia kartkę z listą błędów na stanowisku grupy, której model testowała, i wraca na swoje stanowisko. Tam, w kolejnym ćwiczeniu, poprawi własny model na podstawie listy pozostawionej przez inną grupę. Każda grupa powinna stworzyć swój model, przetestować model innej	
--	--	--

	grupy, dać informację zwrotną innej grupie odnośnie jej modelu (pozostawiona kartka) oraz otrzymać informację zwrotną po powrocie na swoje stanowisko (znaleziona przy stanowisku kartka).	
30 min	Teoria błędów w danych treningowych i poprawa modelu Powiedz: <i>Teraz przez około 15 minut przejdziemy szybko przez najczęstsze błędy, które psują modele AI, do każdego pokażę krótki przykład lub demonstrację. Potem dostaniecie 15 minut na poprawienie własnego modelu na podstawie tego, czego się dowiecie, i listy błędów, którą przyniesiecie od innej grupy. Sprawdzamy razem, czy poprawka pomogła.</i> Ten blok ma więc dwie wyraźne części: najpierw (ok. 15 min) prezentacja dziewięciu błędów w oparciu o przygotowaną wcześniej prezentację lub karty (patrz: Przygotowanie dla trenera), bez przerywania na poprawki po każdym błędzie. Następnie (ok. 15 min) wspólny czas na poprawki: uczestnicy wracają do swoich modeli i stosują wybrane poprawki naraz, korzystając też z listy błędów otrzymanej od innej grupy w ćwiczeniu „Testerzy AI”. Część poprawek warto połączyć w jednym podejściu do fotografowania – np. różne osoby trzymające obiekt i różne odległości od kamery można uzupełnić w tej samej serii zdjęć, co realnie skraca czas potrzebny na poprawki. Błąd 1: Za mało danych, zbyt mały zbiór treningowy Zapytaj grupę: Ile zdjęć zebraliście na daną klasę? Wystuchaj odpowiedzi. Następnie wytłumacz: <i>Model uczy się statystyki – szuka powtarzających się wzorców. Jeśli widzi tylko kilkanaście przykładów jednej klasy, nie uczy się</i>	Rekwizyty do demonstracji Różne tła (biała kartka, ciemna kurtka) Prezentacja błędów (patrz: Przygotowanie dla trenera) <i>*Bloki można skrócić do 4-5 błędów jeśli brakuje czasu. Błędy 1-4 są najważniejsze.</i>

cechy ogólnej, dostawnie zapamiętuje te kilkanaście zdjęć. Na nowych przykładach, których nie widział podczas treningu, po prostu zgaduje. Im więcej zdjęć, tym bardziej ogólne i trafne są wzorce, których model się nauczył. Minimum to około 40–50 zdjęć na klasę, ale naprawdę dobrze jest mieć ich kilka razy więcej.

Demo (opcjonalne): jeśli jest czas, wytrenuj pokazowy model z 5 zdjęciami per klasa i pokaż wyniki na rzutniku. Następnie z 60 zdjęciami, porównajcie procent pewności i stabilność wskazania.

Poproś grupy, które zebrały mniej niż 40 zdjęć na klasę, by zapamiętały to na czas wspólnego bloku poprawek (15 min na końcu tego bloku, patrz wyżej).

Błąd 2: Brak klasy neutralnej

Zapytaj grupę: Co się dzieje z waszym modelem, gdy odsuniesz obiekt od kamery – albo schowasz go za plecami? Daj chwilę na sprawdzenie. Wystuchaj obserwacji. Następnie wytłumacz:

Klasyfikator musi przypisać każde wejście do jednej z klas, to jest jego jedyna opcja. Jeśli nie ma przed kamerą ani banana, ani pomarańczy, model i tak wybierze jedną z dwóch klas, bo nie ma innej możliwości. Bez klasy neutralnej model stale widzi jakiś obiekt, nawet gdy go nie ma. To jest poważny problem w prawdziwych zastosowaniach – wyobraźcie sobie system bezpieczeństwa, który cały czas wykrywa zagrożenie, nawet gdy nikogo nie ma w pomieszczeniu.

Demo: odsuń obiekt od kamery – model pokazuje pewność 60–70% dla jednej z klas. Pokaż to na rzutniku, żeby cała grupa zobaczyła ten efekt.

Poproś grupy o dodanie klasy o nazwie *Nic* lub *Neutral*. W tej klasie należy zebrać około 40 zdjęć samego tła – bez żadnego

obiektu, tylko sam stolik, ściana lub to, co jest za uczestnikami. Klasę tę dodadzą i wytrenują model ponownie podczas wspólnego bloku poprawek (15 min na końcu tej części).

Błąd 3: Monotonne tło- model uczy się otoczenia, nie obiektu

Wytlumacz: AI analizuje cały obraz, nie tylko obiekt, który chcemy rozpoznać. Jeśli wszystkie zdjęcia banana zbieraliście przy tym samym oknie lub przy tej samej ścianie, model mógł nauczyć się, że klasa Banan to tak naprawdę to konkretne okno lub kolor ściany w tle. Po przeniesieniu banana na inne miejsce model przestaje go rozpoznawać – bo nauczył się tła, nie banana.

Demo: poproś jedną grupę, żeby spróbowała rozpoznać swój obiekt na tle zupełnie innym niż podczas treningu – np. położyła go na ciemnej kurtce zamiast na jasnym biurku. Pokaż efekt na rzutniku.

Wspomnij już teraz o słynnym przykładzie z mediów, który omówimy szczegółowo za chwilę: dokładnie ten sam problem dotknął projekt AI, która miała odróżniać wilki od psów rasy husky. Model działał świetnie – dopóki ktoś nie sprawdził dlaczego. Zdjęcia wilków były prawie zawsze robione w śniegu, a zdjęcia husky'ów – na łąkach. Model nauczył się rozpoznawać śnieg, nie psy.

Poprawka: zbierajcie zdjęcia w co najmniej trzech różnych miejscach lub przy różnym oświetleniu. Tę poprawkę zastosujecie razem z innymi w 15-minutowym bloku poprawek na końcu tej części.

Błąd 4: Jeden kąt i jedna odległość fotografowania

Zapytaj: Co się dzieje, gdy obróćcie banana o 90 stopni? A gdy odsuniecie go dalej od kamery? Daj chwilę na sprawdzenie.

Następnie wytłumacz: *Model wytrenowany wyłącznie na zdjęciach banana fotografowanego z boku nie rozpozna banana trzymanego pionowo ani banana z odległości pół metra. AI uczy się cech charakterystycznych dla konkretnego kąta i odległości, nie ogólnej formy obiektu. Wyobraźcie sobie system rozpoznający twarze, który był trenowany wyłącznie na zdjęciach patrzących prosto w kamerę – nie rozpozna twarzy odwróconej o 30 stopni ani twarzy z profilu.*

Poprawka: fotografujcie obiekt ze wszystkich stron, z bliska i z daleka, pod różnymi kątami. Tę serię zdjęć możecie połączyć z poprawką dotyczącą różnych osób (Błąd 5) – fotografujcie z różnych kątów i odległości od razu z udziałem kilku osób z grupy. Wszystko w jednym 15-minutowym bloku poprawek na końcu tej części.

Błąd 5: Brak różnorodności osób (dotyczy modeli rozpoznających ludzi)

Zapytaj grupę: Czy model jednej osoby działał tak samo na ręce innej osoby? Wystuchaj odpowiedzi. Następnie wytłumacz: *Jeśli model był trenowany na zdjęciach rąk tylko jednej osoby – o konkretnej karnacji, rozmiarze dłoni, z określonym manicure – model może słabo działać na rękach innych uczestników. Nie dlatego, że jest zepsuty, ale dlatego, że nigdy nie widział innych przykładów. Ten sam problem w znacznie większej skali dotknął komercyjne systemy rozpoznawania twarzy: badania MIT z 2018 roku wykazały, że wskaźnik błędów dla ciemnoskórych kobiet był kilkadziesiąt razy wyższy niż dla jasnoskórych mężczyzn, ponieważ dane treningowe były zdominowane przez tę drugą grupę. Doprowadziło to do kilku udokumentowanych przypadków*

aresztowania niewinnych osób w USA na podstawie błędnej identyfikacji AI.

Poprawka: jeśli model dotyczy ludzi, zbieraj dane od różnych osób z grupy. Połączcie to z fotografowaniem pod różnymi kątami i odległościami (Błąd 4) – jedna seria zdjęć z udziałem kilku osób załatwi od razu dwie poprawki. Czas na to znajdzie się we wspólnym 15-minutowym bloku poprawek.

Błąd 6: Fałszywa korelacja – model uczy się kontekstu, nie przyczyny

Wytłumacz: AI szuka wzorców statystycznych w całym obrazie. Może przypadkowo nauczyć się rozpoznawać element, który stale towarzyszył obiektowi podczas zbierania danych, zamiast samego obiektu. Klasyczny przykład: AI do wykrywania czerniaka skóry nauczyła się, że obecność linijki na zdjęciu koreluje z niebezpieczną zmianą skórą. Lekarze używali linijki do pokazania skali tylko przy podejrzanych, poważnych zmianach, zwykłe, niegroźne zmiany nie miały linijki na zdjęciu. Model nauczył się: linijka na zdjęciu = niebezpieczna zmiana, choć oczywiście linijka nie ma nic wspólnego z chorobą.

Zapytaj grupę: Sprawdźcie – czy wasz model rozpoznaje wasz obiekt, czy może coś innego, co jest stale widoczne na waszych zdjęciach? Spróbujcie trzymać obiekt na innym tle niż podczas treningu.

Poprawka: upewnij się, że w tle zdjęć nie ma stałego elementu. Zmieniaj tło, oświetlenie, ubranie – a następnie sprawdź, czy model nadal działa.

Błąd 7: Zbyt podobne klasy wizualnie

Wytłumacz: *Jeśli klasy są bardzo do siebie podobne wizualnie – na przykład gest dwóch palców w górę i trzech palców w górę – model potrzebuje dużo więcej przykładów i dużo większego zróżnicowania, żeby wychwycić różnicę. Przy małej liczbie zdjęć i podobnych klasach model będzie się mylił bardzo często, nie dlatego, że jest zepsuty, ale dlatego, że zadanie jest po prostu trudne. AI nie wie, że dla nas różnica między dwoma a trzema palcami jest oczywista – ona musi się tego nauczyć z danych, a to wymaga setek starannie zebranych przykładów.*

Zapytaj grupę: Jak wizualnie różne są wasze klasy? Czy są to rzeczy bardzo różniące się kształtem i kolorem, czy podobne? Wysłuchaj odpowiedzi.

Poprawka: zacznij od klas wyraźnie różnych wizualnie. Gdy model działa stabilnie, możesz go stopniowo komplikować, dodając trudniejsze klasy.

Błąd 8: Nieznany obiekt (brak obsługi „czegoś nowego”)

Zapytaj grupę: Co robi wasz model, gdy pokażecie mu coś zupełnie nowego – coś, czego nie ma w żadnej klasie? Daj chwilę na sprawdzenie. Wysłuchaj odpowiedzi. Następnie wytłumacz:

Klasyfikator nie „wie”, że czegoś nie wie. Jeśli pojawi się nowy obiekt, którego model nigdy nie widział i który nie należy ani do klasy neutralnej, ani do żadnej z wytrenowanych klas, model i tak przypisze go do jednej z istniejących kategorii. Dzieje się tak dlatego, że jego zadaniem jest zawsze wybrać jedną z dostępnych opcji – nawet jeśli żadna nie pasuje.

To jest niebezpieczne w realnych zastosowaniach. Wyobraźcie sobie robota, który zna tylko dwa znaki drogowe – każdy nowy

	znak zostanie przez niego błędnie zinterpretowany jako jeden z tych dwóch. Demo: pokaż modelowi dowolny nowy obiekt (np. kubek, jeśli trenowano na owocach). Zobaczcie, do której klasy zostanie przypisany i z jaką pewnością. Pokaż to na rzutniku. Błąd 9: Niespodziewana cecha deskryptywna (model uczy się jednej dominującej cechy) Wytlumacz: Model czasami nie uczy się tego, co myślimy, że się uczy. Zamiast rozpoznawać cały obiekt, może „zdecydować”, że najłatwiejszą cechą do odróżniania klas jest jedna dominująca właściwość – np. kolor, jasność, albo kształt fragmentu obrazu. Problem polega na tym, że ta cecha może występować także w innych sytuacjach. Wtedy model zaczyna popełniać systematyczne błędy. Przykład: model miał rozpoznawać trzy znaki drogowe, ale zauważył, że czerwony kolor często pojawia się w jednym z nich. Zamiast nauczyć się kształtu znaku, nauczył się koloru. Efekt: • robot zatrzymuje się przed każdym czerwonym obiektem (np. kubkiem), • skręca przy każdym niebieskim obiekcie, nawet jeśli to nie jest znak drogowy. Zapytaj grupę: Czy wasz model reaguje czasem na coś, co nie jest waszym obiektem – ale ma jakąś podobną cechę (np. kolor)? Daj chwilę na sprawdzenie. Demo: pokaż modelowi obiekt o podobnym kolorze, ale zupełnie innym kształcie. Zobaczcie, czy model się „nabierze”.	
--	--	--

	Daj grupom 15 minut na poprawienie swoich modeli. Po omówieniu i poprawieniu wszystkich błędów zapytaj całą grupę: Który z tych błędów był w waszym modelu? Czy poprawka pomogła? Wystuchaj krótkich odpowiedzi od każdej grupy i oceń, które grupy osiągnęły największą poprawę po zastosowaniu korekty.	
10 min	Case'y z mediów – który błąd tu wystąpił? Powiedz: <i>Te błędy, o których rozmawialiśmy, nie są wymysłem – zdarzały się i zdarzają naprawdę w prawdziwych systemach AI używanych przez miliony ludzi. Teraz przejdziemy przez kilka głośnych przykładów z mediów. Waszym zadaniem jest odpowiedzieć: który z siedmiu opisanych błędów doprowadził do tej wpadki? I co należało zrobić inaczej podczas zbierania danych?</i> Case 1: Wilk czy husky? Opowiedz: <i>Pewien projekt naukowy miał na celu stworzenie modelu AI rozróżniającego wilki od psów rasy husky. Model działał świetnie podczas testów – aż ktoś sprawdził dlaczego tak dobrze sobie radzi. Okazało się, że zdjęcia wilków były prawie zawsze robione w śniegu lub na tle zimowego lasu, a zdjęcia husky'ów – na łąkach i trawiastym tle. Model nauczył się rozpoznawać śnieg i zieloną trawę, nie psy. Po tym jak do bazy dodano zimowe zdjęcia husky'ów i letnie zdjęcia wilków, skuteczność modelu drastycznie spadła – okazało się, że naprawdę nie wiedział, jak odróżnić te dwa gatunki, bo nigdy tego nie uczył się z odpowiednich danych.</i>	Wyborcza Biz – wilk vs husky Wikipedia – Tay (bot) MIT – bias w systemach AI Linijka i czerniak (ang.) <i>*Można skrócić do 2–3 case'ów jeśli brakuje czasu.</i> <i>W standardowym wariancie (10 min) omów Case 1-3. Case 4 (TAY) potraktuj jako bonus, jeśli zostanie czas.</i>

Zapytaj grupę: Który błąd z naszej listy tu wystąpił? Wysłuchaj odpowiedzi – szukaj: błąd 3 (monotonne tło) i błąd 6 (fałszywa korelacja). Co należało zrobić inaczej?

Case 2: Rozpoznawanie twarzy i błędne aresztowania

Opowiedz: Komercyjne systemy rozpoznawania twarzy używane przez policję w USA i w innych krajach miały dramatycznie wyższy wskaźnik błędów dla ciemnoskórych osób, szczególnie kobiet.

Badania MIT z 2018 roku wykazały, że najgorsze systemy myliły się w przypadku ciemnoskórych kobiet w 34% przypadków, a dla jasnoskórych mężczyzn wskaźnik błędów wynosił poniżej 1%. W USA odnotowano kilka udokumentowanych przypadków aresztowania niewinnych osób na podstawie błędnej identyfikacji AI. W jednym z nich mężczyzna spędził kilka dni w areszcie, zanim błędne wskazanie zostało zweryfikowane przez policjantów.

Zapytaj grupę: Który błąd z naszej listy? Wysłuchaj odpowiedzi – szukaj: błąd 5 (brak różnorodności demograficznej). Podkreśl, że konsekwencje tego błędu były bardzo poważne dla realnych ludzi.

Case 3: Linijka i czerniak

Opowiedz: Badacze pracowali nad AI mającą wspomagać lekarzy w diagnozowaniu czerniaka skóry na podstawie zdjęć zmian skórnych. Model osiągał imponujące wyniki podczas testów. Aż ktoś zauważył, że model podejrzanie często jako niebezpieczną oznaczał zmianę, gdy na zdjęciu widoczna była linijka. Okazało się, że lekarze używali linijki do pokazania skali i rozmiaru zmiany wyłącznie przy zmianach podejrzanych – tych, które warte były sfotografowania i opisanie. Zwykłe, niegroźne zmiany nie miały linijki na zdjęciu. Model nauczył się: linijka na

zdjęciu = *niebezpieczna zmiana, choć oczywiście linijka nie ma nic wspólnego z chorobą.*

Zapytaj grupę: Który błąd z naszej listy? Wystuchaj odpowiedzi – szukaj: błąd 6 (fałszywa korelacja).

Case 4: Bot TAY – uczenie się od użytkowników bez kontroli

Opowiedz: *W 2016 roku Microsoft wypuścił na Twitterze chatbota o imieniu TAY. Bot miał uczyć się rozmawiania z użytkownikami i stawać się coraz lepszy dzięki interakcjom. W ciągu 24 godzin użytkownicy Twittera celowo trenowali bota, odpowiadając mu w sposób, który chcieli by naśladował – i tak TAY zaczął generować rasistowskie i obraźliwe treści. Microsoft wyłączył bota po niecaterj dobie. Projekt był wielką wpadką wizerunkową i został szeroko omówiony w mediach na całym świecie.*

Zapytaj grupę: Który błąd z naszej listy? Wystuchaj odpowiedzi – szukaj: brak kontroli jakości danych treningowych, brak filtrowania wejść. Podkreśl, że to przykład tego, gdy dane są zbierane bez nadzoru – model uczy się wszystkiego, w tym rzeczy szkodliwych.

Możesz przy tej okazji wprowadzić pojęcia uczenia nadzorowanego i nienadzorowanego. Wasze modele w Teachable Machine uczyły się w sposób nadzorowany – wy sami przypisywaliście zdjęcia do konkretnych, nazwanych klas (np. „banan”, „pomarańcza”), więc model wiedział, jaka jest „poprawna odpowiedź” dla każdego przykładu. TAY uczył się w sposób bliższy nienadzorowanemu – chłonał rozmowy z użytkownikami bez żadnego nadzoru nad tym, co jest treścią właściwą, a co szkodliwą, dlatego nie miał szans odróżnić dobrego wzorca od złego.

	Podsumuj: <i>Jak widzicie, te błędy nie są problemem tylko waszych szkolnych modeli – to są te same błędy, które popełniają wielkie firmy technologiczne dysponujące ogromnymi zasobami. Różnica jest taka, że tam konsekwencje są dużo poważniejsze niż to, że AI myli banana z pomarańczą. Właśnie dlatego wiedza o jakości danych treningowych jest tak ważna.</i>	
*dodatkowe	Informacja o kryteriach oceny dla uczestników Konkursu CDT Warsztat prowadzony jest w ramach projektu konkursowego CDT, poinformuj uczestników o tym, jak wiedza zdobyta dziś przekłada się bezpośrednio na kryteria oceny Etapu I Konkursu. Powiedz: <i>W Etapie I Konkursu CDT wasze zadanie polega na nagraniu krótkiego filmu pokazującego wytrenowany przez was model AI. Jeden z kryteriów oceny dotyczy właśnie opisu procesu trenowania modelu. Jurorzy będą sprawdzać, czy w filmie widać, że rozumiecie, jak działają dane treningowe i jakie błędy mogą się w nich pojawiać. Nie musicie omawiać wszystkich siedmiu błędów – ale im więcej świadomości o jakości danych wykażecie, tym wyższa ocena.</i> Konkretnie jurorzy będą zwracać uwagę na to, czy w filmie widać: - wystarczającą liczbę danych treningowych (kilkadziesiąt zdjęć na klasę, nie kilka) - klasę neutralną, jeśli jest to uzasadnione (model ma opcję „nie rozpoznaję nic”) - zróżnicowanie tła, miejsc i oświetlenia podczas zbierania danych - różne kąty i odległości fotografowania obiektu - różnorodność osób w danych, jeśli model dotyczy ludzi	<i>*Blok informacyjny – omówić ustnie lub rozesłać jako notatkę po zajęciach.</i>

	- świadomość możliwości fałszywej korelacji i sprawdzenie, że model reaguje na sam obiekt, a nie na kontekst - dokumentację procesu iteracyjnego: dane → trening → test → poprawa Podkreśl, że film nie musi być długim dokumentem – najważniejsze jest pokazanie, że uczestnicy świadomie i metodycznie podeszli do tworzenia modelu, a nie tylko kliknęli Train i sprawdzili wynik.	
5 min	Podsumowanie Zapytaj uczestników, czego się dziś dowiedzieli i nauczyli. Przeprowadź krótką rundę zamykającą: każdy uczestnik podaje jedną rzecz, którą zapamięta z tego warsztatu. Zbieraj odpowiedzi swobodnie, bez oceniania. Podkreśl: <i>Uczenie maszynowe opiera się na wzorcach w danych i każdy błąd w danych to potencjalny błąd modelu. Wasz model mylił się nie dlatego, że był zepsuty, tylko dlatego, że dane, na których się uczył, nie były wystarczająco dobre. Tak samo jest z wszystkimi systemami AI na świecie. Różnica między dobrym a złym systemem AI bardzo często leży właśnie w jakości danych, nie w skomplikowanym algorytmie. Dlatego każda osoba, która rozumie ten mechanizm – jak wy po dzisiejszych zajęciach – jest w stanie ocenić i poprawić każdy system AI oparty na uczeniu maszynowym.</i> Podziękuj za uczestnictwo w warsztacie, opowiedz o kolejnych etapach ścieżki AI i zachęć do dalszego udziału w projekcie i konkursie.	–