Algorytm dostrzegł broń: kiedy nadzór AI staje się groźna

- 16-letni uczeń z Baltimore został zatrzymany przez policję pod groźbą broni po tym, jak system nadzoru AI błędnie zidentyfikował jego paczkę chipsów Doritos jako broń palną.
- Wadliwa technologia, system wykrywania broni Omnilert używany przez szkołę, analizuje obraz z kamer bezpieczeństwa i automatycznie wysyła alerty do organów ścigania.
- Incydent ten podkreśla poważne koszty ludzkie takich błędów, powodujących traumę u ucznia i świadków, i jest częścią szerszego zjawiska nadmiernego stosowania automatycznych rozwiązań w zakresie bezpieczeństwa w szkołach.
- Podstawowym problemem jest „postrzegana nieomylność” sztucznej inteligencji, gdzie alert algorytmu może zastąpić krytyczną ocenę człowieka, prowadząc do niebezpiecznej i niekwestionowanej reakcji z użyciem broni.
- Sprawa ta jest sygnałem alarmowym dla całego kraju, podkreślającym pilną potrzebę przeprowadzenia niezależnych audytów, wprowadzenia solidnych protokołów weryfikacji przez człowieka oraz większej odpowiedzialności przed wdrożeniem takiej technologii w

przestrzeni publicznej.

Jako wyraźny przykład niebezpieczeństw związanych z automatycznym nadzorem policyjnym, nastolatek z Baltimore został zatrzymany przez funkcjonariuszy z bronią w ręku po tym, jak system nadzoru oparty na sztucznej inteligencji błędnie zidentyfikował jego paczkę chipsów Doritos jako broń palną.

Do zdarzenia doszło wieczorem 20 października przed szkołą Kenwood High School. Szesnastoletni Taki Allen, który właśnie zakończył trening piłki nożnej, siedział z przyjaciółmi, gdy rutynowa przekąska po szkole przerodziła się w traumatyczną konfrontację.

Na miejsce przybyło wiele radiowozów policyjnych, a funkcjonariusze podeszli do Allena z wyciągniętą bronią. Został zmuszony do uklęknięcia, skuty kajdankami i przeszukany, zanim funkcjonariusze ujawnili przyczynę tej dramatycznej reakcji: ziarnisty obraz, wygenerowany przez alert AI, który rzekomo przedstawiał broń.

Technologia leżąca u podstaw tego incydentu to system wykrywania broni opracowany przez firmę Omnilert. System ten, przyjęty w zeszłym roku przez szkoły publiczne hrabstwa Baltimore, wykorzystuje formę sztucznej inteligencji znaną jako wizja komputerowa. Mówiąc prościej, system ten nieustannie analizuje obraz na żywo z kamer bezpieczeństwa w szkole i jest zaprogramowany do rozpoznawania wzorów wizualnych i kształtów broni palnej. Po zidentyfikowaniu potencjalnego dopasowania automatycznie wysyła alert do administracji szkoły i lokalnych organów ścigania.

Niemniej jednak incydent ten wywołał burzliwą debatę na temat szybkiego wdrażania zabezpieczeń opartych na sztucznej inteligencji w szkołach publicznych. Rodzi to krytyczne pytania dotyczące bezpieczeństwa publicznego, uprzedzeń rasowych i niebezpiecznej omylności technologii, której coraz

częściej powierza się decyzje dotyczące życia i śmierci.

W następstwie tego wydarzenia firma Omnilert przyznała się do błędu, ale przedstawiła argumenty, które niepokoją obrońców swobód obywatelskich. Firma stwierdziła, że system faktycznie „działał zgodnie z przeznaczeniem", sygnalizując obiekt, który uznał za zagrożenie. Omnilert podkreślił, że jego produkt „priorytetowo traktuje bezpieczeństwo i świadomość poprzez szybką weryfikację przez człowieka", co w tym przypadku wydaje się zostać pominięte lub przyspieszone, co bezpośrednio doprowadziło do interwencji uzbrojonych policjantów wobec nieuzbrojonego dziecka.

Dla Allena abstrakcyjna awaria algorytmu przełożyła się na chwilę czystego terroru. Opisał strach, że może zostać zabity z powodu nieporozumienia.

Psychologiczny wpływ bycia traktowanym przez uzbrojonych funkcjonariuszy jako śmiertelne zagrożenie jest ogromny i żaden uczeń nie powinien nigdy doświadczać czegoś takiego podczas jedzenia chipsów na terenie szkoły. Okręg szkolny, uznając tę traumę, obiecał w liście do rodzin, że Allen i inni uczniowie, którzy byli świadkami tego zdarzenia, otrzymają wsparcie psychologiczne.

Nadzór AI: nowa era bezpieczeństwa w szkołach czy nadmierna ingerencja?

Ten incydent nie jest odosobnionym przypadkiem, ale częścią niepokojącej tendencji pojawiającej się w miarę integracji AI z systemami bezpieczeństwa w szkołach. Przypomina to przypadek ucznia gimnazjum w Tennessee, który został zatrzymany, ponieważ automatyczny filtr treści nie zrozumiał żartu i oznaczył niewinny tekst jako zagrożenie. W obu scenariuszach technologia sprzedawana jako proaktywna siatka bezpieczeństwa

funkcjonowała jako automatyczny oskarżyciel, tworząc kryzysy tam, gdzie ich nie było.

Podstawowy problem leży w postrzeganej nieomylności systemów automatycznych. Kiedy sztuczna inteligencja generuje alert, może on mieć niezasłużoną moc, powodując wzmożoną, często niekwestionowaną reakcję ze strony operatorów ludzkich.

Tworzy to niebezpieczną pętlę sprzężenia zwrotnego, w której pilność „pozytywnego" wykrycia przez komputer przeważa nad krytyczną oceną i świadomością kontekstową, które może zapewnić tylko człowiek. Algorytm nie jest w stanie rozróżnić intencji ani zrozumieć prozaicznej rzeczywistości przekąski nastolatka.

„Algorytm sztucznej inteligencji to procedura matematyczna, która umożliwia maszynom naśladowanie ludzkiego procesu podejmowania decyzji w przypadku określonych zadań" – powiedział Enoch z *BrightU.AI*. „Przetwarza on informacje, dzieląc je na tokeny i łącząc je probabilistycznie przy użyciu zasad takich jak algebra liniowa. Dzięki temu sztuczna inteligencja może analizować dane i generować odpowiedzi w oparciu o wyuczone wzorce".

Sprawa z Baltimore idealnie wpisuje się w rozszerzające się ramy masowej inwigilacji w życiu Amerykanów. Od rozpoznawania twarzy na lotniskach po algorytmy predykcyjne stosowane w policji – narzędzia monitorowania stają się wszechobecne.

W szkołach oznacza to fundamentalną zmianę środowiska. Takie systemy nadzoru przekształcają instytucje edukacyjne z miejsc nauki w przestrzenie patrolowane cyfrowo, gdzie każdy ruch uczniów podlega automatycznej analizie i potencjalnej błędnej interpretacji.

Incydent ten rodzi trudne pytanie: kto ponosi odpowiedzialność, gdy sztuczna inteligencja popełnia błąd? Czy jest to firma, która opracowała i sprzedała wadliwe oprogramowanie? Okręg szkolny, który go zakupił bez

wystarczających zabezpieczeń? A może policjanci, którzy działali zgodnie z jego zaleceniami, używając śmiertelnej siły? Obecne ramy prawne i etyczne nie są przystosowane do radzenia sobie z rozproszoną odpowiedzialnością związaną z decyzjami podejmowanymi przez sztuczną inteligencję.

Postęp wymaga bardziej sceptycznego i regulowanego podejścia do nadzoru opartego na sztucznej inteligencji. Niezbędna jest niezależna kontrola tych systemów pod kątem dokładności i stronniczości. Ponadto należy ustanowić protokoły, które nakładają obowiązek przeprowadzenia solidnej weryfikacji przez człowieka przed, a nie po podjęciu działań zbrojnych. Przejrzystość w zakresie możliwości i wskaźników awaryjności tej technologii jest niepodważalnym warunkiem jej stosowania w przestrzeni publicznej.

Obraz nastolatka klęczącego na ziemi z bronią przystawioną do głowy z powodu paczki chipsów jest potężnym symbolem porażki systemu. Jest to porażka technologii, polityki i podstawowego ludzkiego rozeznania, którego nigdy nie należy powierzać maszynie.