Opis i zastosowanie aplikacji FenrirAI do rozpoznawania treści przedstawiających seksualne wykorzystywanie dzieci

Wojciech Oronowicz-Jaśkowiak

Wydział Informatyki Polsko-Japońskiej
Akademii Technik Komputerowych w Warszawie

Treści przedstawiające seksualne wykorzystywanie dzieci (child sexual abuse materials – *CSAM) to zdjęcia, filmy, rysunki lub rendery komputerowe, na których został utrwalony bądź wytworzony wizerunek osoby poniżej 18 r.ż. w czasie kontaktu seksualnego lub prezentujący nagość. Istotną rolę w procesie oceny tych treści odgrywają biegli z zakresu informatyki, antropologii i seksuologii. Z uwagi na rosnącą liczbę zarówno wszczętych postępowań, jak i stwierdzonych przestępstw, które dotyczą uzyskania dostępu, rozpowszechniania lub tworzenia CSAM, są prowadzone badania informatyczne nad opracowaniem narzędzi ułatwiających ocenę tych materiałów. O ile rozwiązania wspierające proces identyfikacji CSAM są dostępne od dłuższego czasu, o tyle w ostatnich latach badania koncentrowały się w szczególności na wykorzystaniu w tym celu uczenia maszynowego (tzw. sztucznej inteligencji) zapewniającego optymalizację procesu ich klasyfikacji. Celem artykułu jest opisanie cyklu publikacji składających się na proces powstawania aplikacji FenrirAI umożliwiających rozpoznanie CSAM, przedstawienie ich zastosowań z perspektywy seksuologicznej i prawnej oraz wskazanie ograniczeń. Omówione zostaną również inne badania związane z zastosowaniem uczenia maszynowego do automatycznej detekcji materiałów pornograficznych i przyszłe kierunki rozwoju oprogramowania. Aplikacje FenrirAI są udostępniane bezpłatnie biegłym sądowym, funkcjonariuszom wymiaru sprawiedliwości i badaczom. Dokumentacja projektu dostępna jest pod numerem DOI: 10.17605/OSF.IO/RU7JX.*

Słowa kluczowe:
treści przedstawiające seksualne wykorzystywanie dzieci, CSAM, informatyka sądowa, uczenie maszynowe

Treści przedstawiające seksualne wykorzystywanie dzieci (*child sexual abuse materials* – CSAM), to materiały graficzne, filmowe lub – rzadziej – pozostałe pliki multimedialne, które przedstawiają osoby poniżej 18 r.ż. w czasie aktywności seksualnej lub prezentujące różny stopień nagości. Problemem, z którym mierzy się polski wymiar sprawiedliwości, jest konieczność przeanalizowania corocznie tysięcy spraw, w których zachodzi podejrzenie uzyskania dostępu do takich materiałów lub ich przechowywania. Zgodnie ze statystykami prowadzonymi przez policję w 2020 r. liczba postępowań wszczętych z art. 202 Kodeksu karnego wynosiła 705, a liczba przestępstw stwierdzonych – 5330 (Policja, 2023).

Z uwagi na charakterystykę CSAM (łączenie dowodu cyfrowego z koniecznością przeprowadzenia oceny antropologicznej i seksuologicznej) materiały te są oceniane przez biegłych z zakresu antropologii i informatyki, co w przypadku konieczności sprawnego sporządzenia opinii może stanowić wyzwanie z uwagi na niewielką liczbę biegłych tych specjalności. Częściowym rozwiązaniem tego problemu jest wykorzystanie aplikacji umożliwiających automatyczną detekcję z wykorzystaniem modeli uczenia maszynowego.

Jeszcze innym problemem systemowym, z którym mierzy się polski wymiar sprawiedliwości, jest – w ocenie autora niniejszego artykułu – brak systematycznej współpracy między badaczami, praktykami wymiaru sprawiedliwości i przedstawicielami organizacji, którzy podejmują działania na rzecz zwalczania CSAM. Taka współpraca mogłaby pozwolić na zwiększenie widoczności podejmowanych działań i opracowanie wspólnych standardów pracy. Ponadto sam system funkcjonowania biegłych w Polsce jest od lat przedmiotem krytyki, np. Najwyższej Izby Kontroli: „Aktualny model funkcjonowania biegłych nie gwarantuje powoływania najlepszych ekspertów terminowo wydających merytoryczne i rzetelne opinie. Brakuje przede wszystkim skutecznej procedury weryfikowania kompetencji kandydatów na biegłych sądowych, a w niektórych specjalnościach także samych kandydatów…" (NIK, 2015). Podkreśla się konieczność zmian systemowych i wprowadzenia specjalistycznych szkoleń dla biegłych, aby działania podejmowane na rzecz ograniczenia dostępu do CSAM były skuteczniejsze.

W ostatnim czasie zaprezentowano w cyklu publikacji naukowych pierwsze polskie aplikacje umożliwiające wykrywanie CSAM z wykorzystaniem sieci neuronowych,

tj. FenrirAI. Celem niniejszego artykułu jest opisanie cyklu badań składających się na opracowanie aplikacji FenrirAI oraz przedstawienie ich zastosowania i bieżących ograniczeń z perspektywy seksuologicznej oraz prawnej.

Przebieg pracy nad materiałami pornograficznymi z udziałem małoletnich

Treści przedstawiające seksualne wykorzystywanie dzieci z uwagi na to, że łączą w sobie aspekty związane z kryminalistyką, seksuologią, antropologią i informatyką sądową[1], są przedmiotem opiniowania interdyscyplinarnego (Szmit i in., 2011). Przedstawiony zostanie zwięźle zarys kompetencji biegłych poszczególnych specjalności.

Proces rozpoznawania i opisywania CSAM rozpoczyna się zazwyczaj od zgłoszenia związanego z nimi incydentu oraz jego dalszej oceny przez funkcjonariuszy policji lub prokuratury. Następnie sprawdzana jest informacja o tym, że dana osoba mogła mieć dostęp do CSAM, posiadała je lub wytwarzała oraz zabezpieczane są nośniki danych. Warto zaznaczyć, że po nowelizacji polskiego Kodeksu karnego karalne jest uzyskiwanie dostępu do treści przedstawiających seksualne wykorzystywanie osób poniżej 18 r.ż. oraz przechowywanie takich treści lub ich wytarzanie. Przed nowelizacją – karalne było uzyskiwanie dostępu do takich treści z udziałem małoletniego rozumianego jako osoba przed ukończeniem 15 r.ż. oraz rozpowszechnianie takich treści lub ich wytwarzanie.

W przypadku uzasadnionego podejrzenia, że dana osoba popełniła przestępstwo związane z CSAM są zabezpieczane nie tylko komputery, ale również płyty CD/DVD, pamięci przenośne (pendrive), telefony komórkowe, aparaty i karty pamięci.

W ramach tzw. triage'u w miejscu przeszukania można wykonać oględziny w obecności specjalisty/biegłego z zakresu informatyki. Triage jest strategią szybkiej i wstępnej oceny źródeł dowodowych z perspektywy ich potencjalnego znaczenia dla sprawy. Wykorzystanie wówczas aplikacji sFenrirAI do przeskanowania całego zabezpieczonego nośnika lub wybranego folderu pozwoliłoby na wstępną oceną materiałów audiowizualnych pod względem obecności CSAM. Analogiczne będzie użycie aplikacji podczas oględzin tzw. wtórnych nośników danych (płyta CD, dysk

1 Szmit (2014) proponuje używanie terminu *informatyka sądowa* zamiast *informatyka śledcza*, zauważając, że pojęcie informatyka śledcza występuje równolegle wśród podmiotów komercyjnych. Warto jednak zauważyć, że ten termin jest obecnie wykorzystywany przez niektóre uczelnie w nazwach studiów podyplomowych, sylabusach przedmiotów czy anglojęzycznych tłumaczeniach podręczników.

twardy, laboratoryjny zasób sieciowy, kopia binarna/logiczna nośnika), na których zgodnie z procedurami zabezpieczono istotne z punktu widzenia dane. W obu przypadkach aplikacja sFenrirAI może pomóc we wskazaniu plików, które z wysokim prawdopodobieństwem będą stanowić CSAM.

Zadaniem biegłego wykorzystującego wiedzę z zakresu informatyki jest bezpieczne wyodrębnienie z danego nośnika materiałów, które mogą stanowić CSAM (Szmit, 2014). Do jego kompetencji należy praca z materiałem dowodowym w taki sposób, aby nie naruszyć integralności danych. Materiał dowodowy zapisany na nośnikach elektronicznych jest podatny na zmianę parametrów plików, co na dalszym etapie sprawy mogłoby doprowadzić do utraty jego wartości w postępowaniu sądowym. Aby temu przeciwdziałać, są wykorzystywane specjalne blokery zapisu danych lub systemy operacyjne, które umożliwiają pracę z danymi w trybie tylko do odczytu. Dalej biegły z zakresu informatyki przywraca pliki usunięte wcześniej przez użytkownika, wykorzystując do tego specjalistyczne oprogramowanie do odzyskiwania danych. Następnie jest analizowana zawartość urządzenia i nośnika danych (jeżeli jest dostępna), taka jak historia przeglądania stron internetowych, pobierane pliki, zakładki stron internetowych, wiadomości itd. Wreszcie są analizowane w sposób ręczny lub częściowo automatyczny zdjęcia i filmy mogące potencjalnie stanowić CSAM. Istotne jest, że biegły z zakresu informatyki nie jest specjalistą mogącym stwierdzić, czy dany materiał zawiera wizerunki osób poniżej 18 r.ż., nie może się również odnosić do tego, czy jest to materiał, który z perspektywy seksuologicznej można wskazać jako CSAM. O ile zadaniem biegłego z zakresu informatyki nie jest zatem samodzielne decydowanie o charakterze zabezpieczonych plików, o tyle do jego kompetencji należy już wyodrębnienie wszystkich potencjalnych zdjęć i filmów mogących stanowić CSAM i różnicowanie ich od wszystkich innych typów materiałów na danym nośniku. W przypadku występowania materiału dowodowego z dużą ilością potencjalnych CSAM zasadne mogłoby być wprowadzenie dodatkowego skanowania dysku w celu ich poszukiwania, aby dążyć do minimalizacji błędów ich przeoczenia. Rozwiązaniem tego problemu mogłoby być zastosowanie aplikacji sFenrirAI.

Zadaniem biegłego z zakresu antropologii jest natomiast wskazanie, na których z zabezpieczonych plików znajdują się wizerunki osób, których cechy wyglądu wskazują na wiek poniżej 18 lat. Do oceny stopnia dojrzałości seksualnej wykorzystywana jest skala Tannera (Marshall i Tanner, 1969, 1970), określana również jako *Skala dojrzałości seksualnej* (*Sexual Maturity Scale* – SMR), która jest często używanym narzędziem do oceny rozwoju fizycznego i dojrzałości płciowej u dzieci i młodzieży. Skala ta opiera się na ocenie różnych cech fizycznych, takich jak rozwój narządów płciowych, owłosienie, rozwój piersi itd. Mimo że SMR może być pomocna w pewnym stopniu

w ocenie antropologicznej, to w praktyce jest ona uzupełniana wiedzą z zakresu antropologii. Ponadto jej zastosowanie w sprawach związanych z CSAM stanowi przedmiot kontrowersji (Rosenbloom, 1998). Warto również dodać, że opiniowanie antropologiczne nie może dostarczyć dokładnych informacji odnoszących się do wieku osób przedstawionych na danych materiałach, a stanowi jedynie systematyczną próbę oszacowanie wieku danej osoby. Wskazane ograniczenie wiąże się z tym, że rozwój cechuje się znacznym zróżnicowaniem międzyosobniczym i jest modyfikowany przez czynniki środowiskowe, takie jak odżywianie, warunki życia, czynniki genetyczne itd. Problemem, który występuje na tym etapie pracy, jest konieczność samodzielnego decydowania o tym, który z zabezpieczonych plików może zawierać wizerunki osób małoletnich, a które z plików zawierają wizerunki osób dojrzałych seksualnie. Istotne byłoby zapewnienie narzędzia, które bazowałoby na dużej liczbie ocenionych już wcześniej antropologicznie i seksuologicznie CSAM i zwracałby informację o rozpoznawanym stopniu rozwoju narządów płciowych wraz z dostarczeniem informacji o uzasadnieniu predykcji. Rozwiązaniem tego problemu również mogłoby być zastosowanie aplikacji xFenrirAI i mFenrirAI.

Podsumowując, opiniowanie CSAM wiąże się z wieloma trudnościami, których rozwiązanie (lub złagodzenie) mogłoby się wiązać z zastosowaniem służących do tego celu aplikacji wykorzystujących uczenie maszynowe.

Uczenie maszynowe

Uczenie maszynowe jest dziedziną informatyki wiążącą się z opracowaniem i rozwojem algorytmów, które umożliwiają uczenie się na podstawie danych i w dalszej kolejności podejmowanie samodzielnych decyzji (Garcia-Garcia i in., 2017; Krizhevsky i in., 2017). Uczenie maszynowe jest często utożsamiane ze „sztuczną inteligencją", jednak z perspektywy informatycznej są to różne procesy. Celem prowadzenia badań nad sztuczną inteligencją jest opracowanie takich algorytmów, które będą możliwie wiernie symulowały procesy poznawcze przede wszystkim ludzkie. Uczenie maszynowe wykorzystuje natomiast takie algorytmy, które „dostosowują się" do danych prezentowanych w procesie treningu. W przeciwieństwie do tradycyjnych algorytmów, w których programista musi ręcznie zaplanować (wszystkie lub większość z nich) kroki rozwiązania danego problemu uczenie maszynowe pozwala na adaptację na podstawie zgromadzonych danych (Goodfellow i in., 2017). Uczenie maszynowe jest często kojarzone z sieciami neuronowymi, które są implementowane z sukcesem w wielu zadaniach polegających na klasyfikacji zdjęć lub detekcji obiektów. Warto jednak dodać, że istnieje wiele innych algorytmów uczenia maszynowego (Hastie i in., 2009),

takich jak regresja, algorytm k-nn (k najbliższych sąsiadów) i drzewa decyzyjne, które również są z powodzeniem stosowane w rozwiązaniach technologicznych wykorzystujących uczenie na podstawie danych. Z uwagi na ich charakterystykę nie są one jednak często wykorzystywane w zadaniach klasyfikacji CSAM.

Z punktu widzenia niniejszego artykułu na szczególną uwagę zasługuje możliwość zastosowania sieci neuronowych w zadaniach klasyfikacji zdjęć (np. różnicowania zdjęć pornograficznych od innych niż pornograficzne) i detekcji (np. w ustaleniu, czy na danym zdjęciu znajdują się narządy płciowe osoby małoletniej poniżej 18 r.ż.; Deng i in., 2009; He i in., 2016; LeCun i in., 2015). Uczenie maszynowe jest niezwykle wydajnym podejściem przy klasyfikacji i detekcji zdjęć ze względu na zdolność do automatycznego wyodrębniania cech i wzorców z obrazów. Tradycyjne metody klasyfikacji obrazów opierają się przede wszystkim na manualnym zdefiniowaniu reguł i cech, które mają być rozpoznawane, co jest czasochłonne. Istnieją również metody identyfikacji zdjęć, które bazują na charakterystykach samego zdjęcia (np. są związane z analizą histogramów), jednak to sieci neuronowe zapewniają wysoką dokładność klasyfikacji. Za pomocą głębokich sieci neuronowych jest możliwe zbudowanie złożonych modeli, które wykorzystują wiele funkcyjnych warstw sieci do wydajnej analizy obrazów. Prezentowane dalej rozwiązanie technologiczne FenrirAI bazuje właśnie na głębokich sieciach neuronowych.

Przebieg badań nad rozwojem aplikacji FenrirAI

Badania prowadzące do opracowania aplikacji umożliwiających rozpoznawanie CSAM były prowadzone w latach 2018–2023 na Uniwersytecie Warszawskim. Niektóre zadania badawcze zostały współfinansowane ze środków budżetowych jako projekt badawczy w ramach programu Inicjatywa Doskonałości – Uczelnia Badawcza Uniwersytetu Warszawskiego: IDUB-622-33/2022 „Środowiska uruchomieniowe modeli uczenia maszynowego w identyfikacji treści przedstawiających seksualne wykorzystywanie dzieci" oraz IDUB-SOB/D110/2023 „Cykl warsztatów prezentujący wyniki badań nad stworzeniem aplikacji do automatycznego rozpoznawania treści przedstawiających seksualne wykorzystywanie dzieci z wykorzystaniem sieci neuronowych".

Z uwagi na pracę ze szczególnym rodzajem materiału wdrożono odpowiednie procedury bezpieczeństwa. Przed rozpoczęciem badań uzyskano zgody lokalnej komisji etycznej i organów prowadzących dane postępowania karne na wykorzystanie materiału do celów badawczych. Wdrożono dodatkowe zabezpieczenia związane z minimalizacją ryzyka podwójnego wykorzystania wyników badań. Materiał

treningowy użyty do właściwych badań informatycznych był wcześniej oceniany z wykorzystaniem ocen antropologicznej i seksuologicznej.

Przed właściwymi badaniami prowadzono prace przygotowawcze obejmujące przegląd wybranych zagadnień seksuologii sądowej i uczenia maszynowego. Przeprowadzono wywiady technologiczne z funkcjonariuszami policji i przedstawicielami prokuratury. Ich celem było uzyskanie wiedzy o najbardziej pożądanych cechach tworzonego rozwiązania technologicznego, dostosowanie go do oczekiwań przyszłych użytkowników i poznanie ograniczeń jego zastosowania. Prezentowano również hipotetyczny zakres zastosowania rozwiązań z zakresu uczenia maszynowego w seksuologii sądowej (Oronowicz-Jaśkowiak, 2019a). Pierwsze badania technologiczne prowadzono nad klasyfikacją treści pornograficznych z udziałem dorosłych i porównywano kilka standardowych architektur sieci neuronowych (Oronowicz-Jaśkowiak, 2019b; Oronowicz-Jaśkowiak i in., 2020; Oronowicz-Jaśkowiak i in., 2022; Oronowicz-Jaśkowiak i Wasilewski, 2022). Dalsze badania przygotowawcze koncentrowały się na możliwie najlepszym dostrojeniu hiperparametrów augmentacji danych (tj. jednego z etapów uczenia maszynowego, który wpływa na proces treningu sieci; Oronowicz-Jaśkowiak, 2021). Właściwe badania zmierzające do opracowania modelu umożliwiającego identyfikację CSAM obejmowały porównywanie kilkuset modeli uczenia maszynowego i dobór optymalnej architektury przy zapewnieniu odpowiednich warunków trenowania sieci. Zapewniono kilka modeli uczenia maszynowego, które zostały przygotowane do klasyfikacji kluczowych kategorii zdjęć z punktu widzenia identyfikacji CSAM. Przygotowano również model detekcji obiektów anatomicznych. Zgodnie z wytycznymi Unii Europejskiej w zakresie godnej zaufania sztucznej inteligencji (Komisja Europejska, 2021) wszystkie etapy pracy nad stworzeniem sieci zostały udokumentowane, ponadto prowadzono prace w obszarze możliwości zapewnienia wyjaśnialności klasyfikacji, co stanowi jeden z elementów wykorzystania uczenia maszynowego w obszarach podwyższonego ryzyka. Predykcje dokonywane przez model weryfikowano zgodnie z teorią antropologiczną (Oronowicz-Jaśkowiak, 2022). Ponadto weryfikowano podatność rozwiązania technologicznego na określone zbiory danych.

W dalszym ciągu są prowadzone prace nad ulepszeniem aplikacji oraz wprowadzeniem nowych funkcjonalności sugerowanych przez funkcjonariuszy policji i biegłych sądowych. Planowane jest wprowadzenie zaawansowanych raportów dokonywanych predykcji, zapewnienie możliwości klasyfikacji różnych rodzajów plików i dalsze dążenie do zwiększenia dokładności klasyfikacji.

Warto dodać, że rozwiązanie technologiczne FenrirAI nie jest jedynym oprogramowaniem umożliwiającym dokonanie automatycznej klasyfikacji CASM. Przed

wykorzystywaniem w tym celu modeli uczenia maszynowego wprowadzano również rozwiązania bazujące na innych technologiach, np. była to analiza histogramów (tj. kolorystyki zdjęcia; Laranjeira da Silva i in., 2022). Dostępne są programy udostępniane bezpłatnie funkcjonariuszom wymiaru sprawiedliwości (np. NuDetective; de Castro Polastro i da Silva Eleuterio, 2010), oprogramowanie komercyjne (np. Media Detective [2018], Snitch Plus [2017]), prowadzone są inne prace badawcze (np. Lin i in., 2003; Santos i in., 2012; Setyanto i in., 2019; Vitorino i in., 2018), ponadto twórcy aplikacji mają możliwość skorzystać z systemów detekcji treści pornograficznych implementowanych do wybranych usług. Zastosowanie sieci neuronowych w zadaniu polegającym na identyfikacji treści pornograficznych zostało wcześniej przeprowadzone w badaniach Moustafy (2015). Na szczególną uwagę zasługują badania zespołu Vitorino i in. (2018), gdzie zaprezentowano model „2-tiered SEIC Detector – ext", który cechował się relatywnie wysoką (86,5%) dokładnością klasyfikacji treści przedstawiających seksualne wykorzystywanie dzieci. Prezentowane są również takie rozwiązania technologiczne, które umożliwiają zapewnienie dodatkowej warstwy ochrony przed aktywnością użytkowników mediów społecznościowych, polegające na blokowaniu CSAM w internecie (Google, 2023). Poza rozwojem aplikacji FenrirAI w Polsce są podejmowane również istotne działania zmierzające do wprowadzenia technologii informacyjnych umożliwiających identyfikację CSAM (APAKT, 2020; SPOX, 2021).

Warto ponadto odnieść się do kwestii etycznych związanych z projektowaniem, wdrażaniem i utrzymaniem aplikacji umożliwiających wykrywanie CSAM. Nie ulega wątpliwości, że podejmowanie wymienionych działań należy do pracy wysokospecjalistycznej, która generuje wysokie koszty. Wątpliwości etyczne i prawne mogą budzić działania podmiotów komercyjnych polegające na uzyskaniu dostępu do CSAM i wykorzystanie ich jako bazy treningowej do wytrenowania własnego modelu uczenia maszynowego sprzedawanego dalej w ramach aplikacji. Autor niniejszego artykułu uważa, że rozwój aplikacji umożliwiających wykrywanie CSAM powinien być finansowany ze środków budżetowych w ramach stypendiów, realizacji badań poprzez finansowanie w drodze konkursów projektów badawczych. Efekty realizacji tych działań pozostałyby wtedy bezpłatne dla wszystkich zainteresowanych grup docelowych narzędzia, a także wprowadzona zostałaby kontrola merytoryczna nad prowadzonymi działaniami.

Dotychczas wykorzystywane w Polsce oprogramowanie cechowało się ograniczeniami. Przede wszystkim nie było dostosowane do nowych regulacji Unii Europejskiej w zakresie stosowania oprogramowania wykorzystującego uczenie maszynowe. Zgodnie z nowymi wytycznymi użycie sztucznej inteligencji w obszarach podwyższonego ryzyka (np. wymiaru sprawiedliwości lub zdrowia) wymaga

dostosowania się do wielu wytycznych. Dotychczas prezentowane rozwiązania nie były projektowane z tej perspektywy. Rozwiązania komercyjne często nie są przeznaczone do klasyfikacji treści CSAM, tj. w ich wypadku dokładność klasyfikacji jest niższa, a ponadto oprogramowanie wykrywa materiały pornograficzne z udziałem osób po osiągnięciu dojrzałości seksualnej, co nie stanowi naruszenia norm prawnych (jeżeli dodatkowo nie jest stosowana przemoc lub nie są wykorzystywane zwierzęta). Nie prowadzono wywiadów technologicznych (z wyjątkiem badań Sanchez i in., 2019). Rozwiązania te nie zapewniały również odpowiednio wysokiej dokładności klasyfikacji. W dotychczas prezentowanych badaniach dokładność klasyfikacji CSAM wynosiła 86,5%. Dokładność taką trudno jest uznać za wystarczającą do zastosowania oprogramowania w praktyce wymiaru sprawiedliwości. Dodatkowym ograniczeniem wybranych rozwiązań jest również ich wysoki koszt i konieczność zakupu przez jednostki policji lub biegłych sądowych.

Aplikacja sFenrirAI

Aplikacja sFenrirAI, która jest przeznaczona na urządzenia z systemem Windows, wykorzystuje modele uczenia maszynowego do automatycznego oznaczania i klasyfikowania zdjęć mogących stanowić CSAM. Jej działanie przypomina pracę programu antywirusowego i polega na skanowaniu wybranego nośnika danych i w przypadku wykrycia zdjęć mogących stanowić CSAM kopiowaniu ich do wybranego katalogu (rys. 1). Funkcjonalność aplikacji przedstawiono na rys. 1. Jeżeli aplikacja rozpozna, z zadanym stopień prawdopodobieństwa materiały CSAM (kategoria 1.) lub materiały pornograficzne z udziałem osób dorosłych (kategoria 2.; celem dodatkowej weryfikacji), to zostaną one skopiowane i wygenerowany zostanie raport. W tabeli 1 przedstawiono zalety i ograniczenia narzędzia.

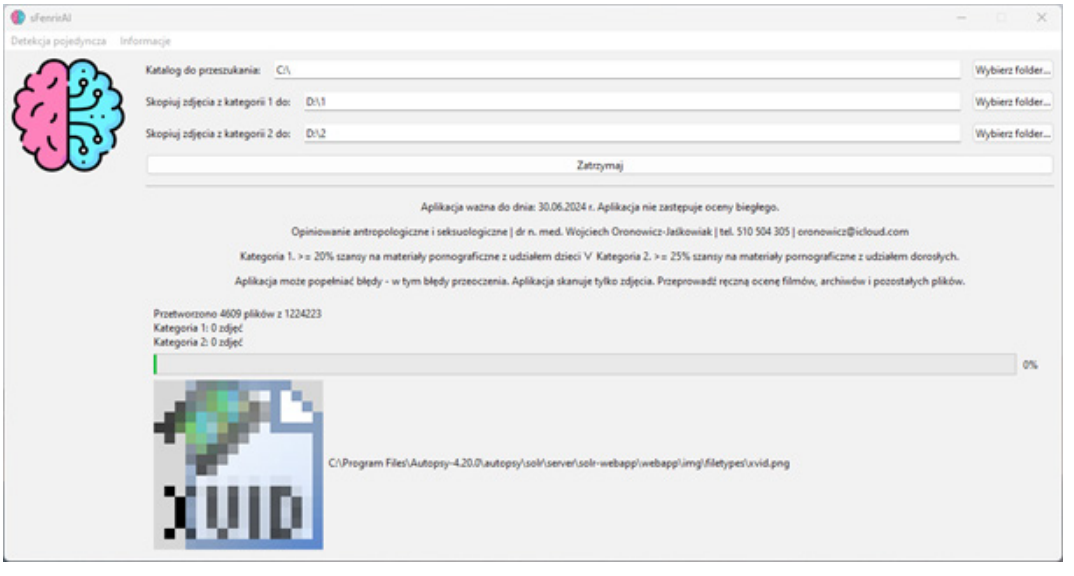
Tabela 1

Zalety i ograniczenia aplikacji sFenrirAI w bieżącej wersji 1.7

Zalety	Ograniczenia
– możliwość przeskanowania całego nośnika danych – możliwość przeskanowania różnych nośników danych (płyty CD/DVD, pamięci przenośne, dyski twarde itd.) – brak konieczności wskazania poszczególnych pod folderów do skanowania – wykrywanie treści również w folderach „ukrytych" – generowanie raportu w postaci plików tekstowych wskazujących wykryte pliki i ich ścieżki	– aplikacja nie skanuje filmów ani innych plików niż graficzne (np. archiwów) – aplikacja tylko na komputery z systemem operacyjnym Windows, brak wersji na komputery z systemem macOS – brak implementacji algorytmu grad-CAM – wykorzystanie tylko jednej sieci neuronowej (związane z potrzebą zapewnienia odpowiedniego czasu predykcji) – brak implementacji detekcji obiektów – brak implementacji algorytmu grad-CAM

Rysunek 1

Zrzut ekranu programu sFenrirAI podczas przeszukiwania wybranego przez użytkownika katalogu (C:\). W katalogu tym zidentyfikowano 1 224 223 pliki. Przeszukano 4609 plików, nie rozpoznano treści CSAM (kategoria 1.) ani materiałów pornograficznych z udziałem osób dorosłych (kategoria 2.). Na dole aplikacji wyświetlana jest ikona zdjęcia, które aktualnie podlega ocenie przez sieć, i ścieżka dostępu, która wskazuje miejsce zapisu pliku.



Aplikacja mFenrirAI

Aplikacja mFenrirAI jest przeznaczona na urządzenia firmy Apple z systemem iOS (iPhone i iPad). Za jej pomocą można wykonać zdjęcie z aparatu urządzenia lub wybrać je z biblioteki i dokonać jego klasyfikacji. Można również wykorzystać aparat urządzenia i dokonać detekcji obiektów znajdujących się na danym materiale, np. uzyskać informację o szacowanym stopniu rozwoju cech antropologicznych i seksuologicznych twarzy, sylwetki, narządów płciowych itd. (rys. 2). Z uwagi na kwestie związane z bezpieczeństwem wykorzystania aplikacji (Hayran i in., 2016) nie jest obecnie planowane wydanie aplikacji mFenrirAI na urządzenia z systemem Android. W tabeli 2 przedstawiono zalety i ograniczenia narzędzia.

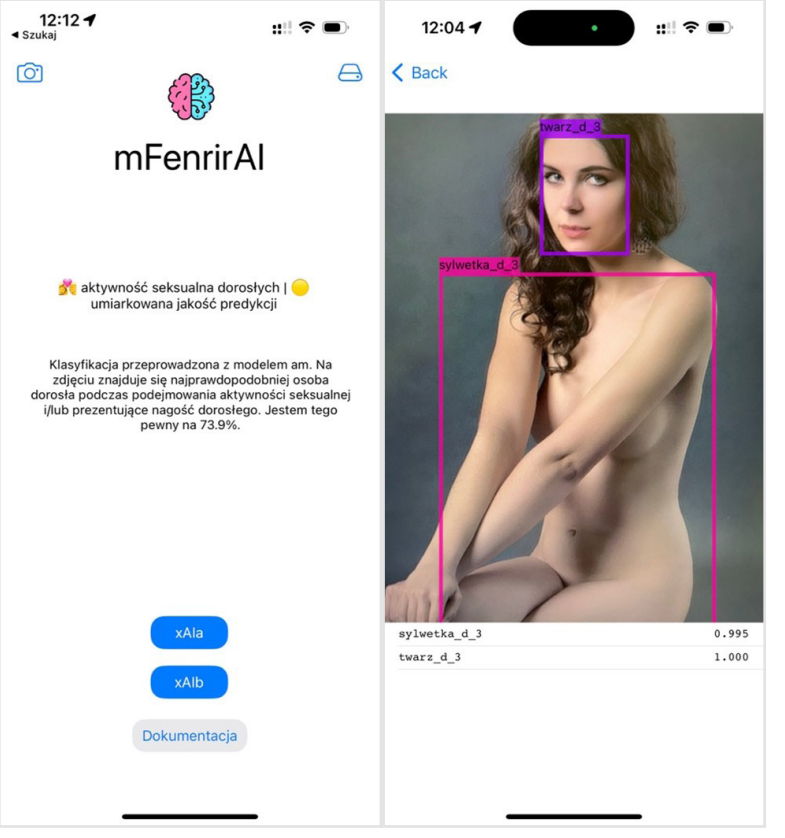
Tabela 2

Zalety i ograniczenia aplikacji mFenrirAI w bieżącej wersji 2.7.1

Zalety	Ograniczenia
– możliwość dokonania detekcji obiektów anatomicznych widocznych na zarówno zdjęciach, jak i materiałach wideo – możliwość dokonania oceny zdjęć występujących w postaci innej niż elektroniczna (np. materiały drukowane z CSAM), – szybka predykcja materiału – możliwość dokonania oceny materiału z wykorzystaniem aparatu urządzenia i znajdującego się w pamięci urządzenia – wykorzystanie dwóch równoległych sieci neuronowych	– aplikacja tylko na urządzenia z systemem operacyjnym iOS, brak wersji na urządzenia z systemem Android – brak implementacji algorytmu grad-CAM

Rysunek 2

Zrzut ekranu aplikacji mFenrirAI. Aplikacja rozpoznała treści pornograficzne z udziałem osób dorosłych i wskazała prawdopodobieństwo predykcji. Następnie z wykorzystaniem modułu detekcji obiektów zostały wskazane cechy antropologiczne ocenianego zdjęcia, tj. twarz i cechy sylwetki wskazujące na osiągnięcie dojrzałości seksualnej. Źródło pochodzenia zdjęcia: www.pixabay.com, domena publiczna.



Aplikacja xFenrirAI

Aplikacja xFenrirAI została przygotowana na urządzenia z systemem Linux. Możliwe jest również jej uruchomienie aplikacji na komputerze z systemem operacyjnym Windows z wykorzystaniem technologii *Windows Subsystem for Linux*. Aplikacja umożliwia klasyfikację i detekcję wybranych cech anatomicznych. Zaimplementowano w niej zmodyfikowany algorytm grad-CAM (Selvaraju i in., 2017; Zhou i in. 2016), co umożliwia poznanie przyczyn dokonania określonej klasyfikacji. Na rys. 3 przedstawiono predykcję dokonaną z wykorzystaniem aplikacji i mapę ciepła z wykorzystaniem algorytmu grad-CAM, która wskazuje obszary zdjęcia mające największe znaczenie dla sieci przy dokonywaniu predykcji. W tabeli 3 przedstawiono zalety i ograniczenia narzędzia.

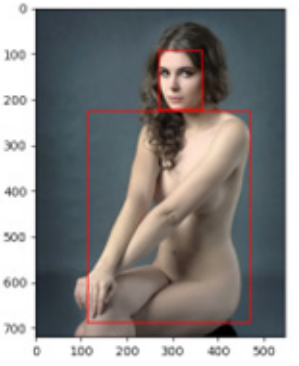
Tabela 3

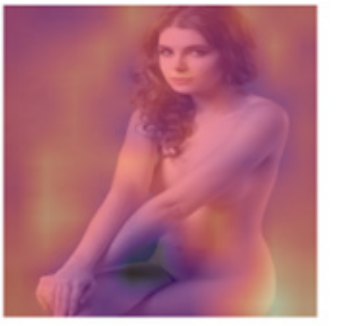
Zalety i ograniczenia aplikacji xFenrirAI w bieżącej wersji 1.0

Zalety	Ograniczenia
– możliwość detekcji obiektów anatomicznych widocznych na zdjęciach – implementacja algorytmu grad-CAM – wykorzystanie kilku równoległych sieci neuronowych	– aplikacja tylko na urządzenia z systemem operacyjnym Linux, brak wersji na urządzenia z systemem Windows, – wydłużony czas predykcji materiałów (w porównaniu z pozostałymi aplikacjami) – aplikacja nie skanuje filmów ani innych plików niż graficzne (np. archiwów)

Rysunek 3

Zrzut ekranu aplikacji xFenrirAI. Aplikacja rozpoznała treści pornograficzne z udziałem osoby dorosłej, stwierdziła, że zdjęcie nie jest miniaturą i jest wyraźne oraz określiła klasę kolorystyki zdjęcia (wykorzystywaną dalej przy dynamicznej modyfikacji algorytmu grad-CAM). Rozpoznała także sylwetkę i twarz charakterystyczną dla kobiet po osiągnięciu dojrzałości seksualnej. Widoczne jest, że istotnym obszarem przy klasyfikacji tego zdjęcia są uda i podbrzusze. Źródło pochodzenia zdjęcia: www.pixabay.com, domena publiczna.





Jak uzyskać dostęp do aplikacji?

Aplikacje FenrirAI są udostępniane bezpłatnie biegłym sądowym, funkcjonariuszom wymiaru sprawiedliwości i badaczom. Aby uzyskać do nich dostęp, należy wysłać wiadomość na adres oronowiczjaskowiak@pjwstk.edu.pl. Informacje dotyczące rozwoju i aktualizacji oprogramowania są publikowane na stronie internetowej: www.oronowicz-jaśkowiak.pl (zakładka software). Dokumentacja projektu dostępna jest pod numerem DOI: 10.17605/OSF.IO/RU7JX.

E-mail autora: **oronowiczjaskowiak@pjwstk.edu.pl**.

Bibliografia

APAKT. (2020). Strona internetowa NASK. https://www.nask.pl.

de Castro Polastro, M., da Silva Eleuterio, P. M. (2010). *Nudetective: A forensic tool to help combat child pornography through automatic nudity detection*. IEE Workshops on Database and Expert Systems Applications.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. W: *2009 IEEE conference on computer vision and pattern recognition* (s. 248–255). Ieee.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857.

Goodfellow, I., Bengio, Y., Courville, A. (2017). *Deep learning (adaptive computation and machine learning series)*. Cambridge Massachusetts.

Google. (2023). *Fighting child sexual abuse online*. https://www.protectingchildren.google.com

Hastie, T., Tibshirani, R., Friedman, J. H., Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hayran, A., İğdeli, M., Yilmaz, A., Gemci, C. (2016). Security evaluation of IOS and Android. *International Journal of Applied Mathematics Electronics and Computers*, 258–261. http://dx.doi.org/10.18100/ijamec.270378

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. W: *Proceedings of the IEEE conference on computer vision and pattern recognition* (s. 770–778).

Komisja Europejska. (2021). Artificial Intelligence Act. Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence and amending certain union legislative acts. http://www.eur-lex.europa.eu/

Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.

Laranjeira da Silva, C., Macedo, J., Avila, S., dos Santos, J. (2022). *Seeing without looking: Analysis pipeline for child sexual abuse datasets*. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency.

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. https://doi.org/10.1038/nature14539

Lin, Y. C., Tseng, H. W., Fuh, C. S. (2003). *Pornography detection using support vector machine*. W: 16th IPPR Conference on Computer Vision, Graphics and Image Processing (s. 123–130).

Marshall, W. A., Tanner, J. M. (1969). Variations in pattern of pubertal changes in girls. *Archives of Disease in Childhood*, *44*(235), 291–303. https://doi.org/10.1136%2Fadc.44.235.291

Marshall, W. A., Tanner, J. M. (1970). Variations in the pattern of pubertal changes in boys. *Archives of Disease in Childhood*, *45*(239), 13–23. https://doi.org/10.1136%2Fadc.45.239.13

Media Detective. (2018). *Strona internetowa Media Detective*. https://www.mediadetective.com/

Moustafa, M. (2015). *Applying Deep Learning to Classify Pornographic Images and Videos*. arXiv preprint arXiv:1511.08899.

NIK. (2015). *Funkcjonowanie biegłych w wymiarze sprawiedliwości*. https://www.nik.gov.pl/plik/id,9608,vp,11856.pdf

Oronowicz-Jaśkowiak, W. (2019a). The application of neural networks in the work of forensic experts in child abuse cases. *Advances in Psychiatry and Neurology*, *28*(4), 273–282.

Oronowicz-Jaśkowiak, W. (2019b). Classification of seven types of legal pornography using a neural network. *Sexological Review*, *19*(1), 32–40.

Oronowicz-Jaśkowiak, W. (2021). *Weryfikacja empiryczna wartości domyślnych augmentacji danych biblioteki fastai* [praca magisterska-inżynierska, Wydział Informatyki Polsko-Japońskiej Akademii Technik Komputerowych w Warszawie]. http://dx.doi.org/10.13140/RG.2.2.28464.28165

Oronowicz-Jaśkowiak, W. (2022). *Analiza antropologiczna trafności klasyfikacji dokonywanej przez sieci neuronowe między okresami dojrzewania dzieci przedstawionymi na materiałach pornograficznych* [praca magisterska, Wydział Biologii Uniwersytetu Warszawskiego].

Oronowicz-Jaśkowiak, W., Bzikowska, E., Jabłońska, K., Kłok, A. (2020). Binary classification of pornographic and non-pornographic materials using the sAI 0.4 model and the modified database. *Advances in Psychiatry and Neurology*, *29*(2), 108–119.

Oronowicz-Jaśkowiak, W., Wasilewski, P. (2022). Description of the neural network based on AB/DL pictures. Possible implications for forensic sexology. Advances in Psychiatry and Neurology, *31*(4), 1–6.

Oronowicz-Jaśkowiak., W., Siwiak, A., Róg, K., Oronowicz-Jaśkowiak, A. (2022). Classification of nine types of pornographic materials using the sAI 0.3 model. *Psychiatria Polska*, *56*(4), 877–888.

Policja. (2023). Strona internetowa Policji. https://statystyka.policja.pl/

Rosenbloom, A. L. (1998). Misuse of Tanner puberty stages to estimate chronologic age. *Pediatrics*, *102*(6), 1494–1494. https://doi.org/10.1542/peds.102.6.1494

Sanchez, L., Grajeda, C., Baggili, I., Hall, C. (2019). A practitioner survey exploring the value of forensic tools, ai, filtering & safer presentation for investigating child sexual abuse material (CSAM). *Digital Investigation*, *29*, 124–142. https://doi.org/10.1016/j.diin.2019.04.005

Santos, C., dos Santos, E. M., Souto, E. (2012). *Nudity Detection Based on Image Zoning*. 11th International Conference on Information Science, Signal Processing, and their Applications (ISSPA).

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017). *Grad-cam: Visual Explanations from Deep Networks via Gradient-based Localization*. Proceedings of the IEEE International Conference on Computer Vision.

Setyanto, A., Kusrini, K., Agastya, I. M. A. (2019). *Comparison of SIFT and SURF Methods for Porn Image Detection*. 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE).

Snitch Plus. (2017). *Strona internetowa Hyperdyne Software*. https://www.hyperdyne-software.com/

SPOX. (2021). Strona Internetowa Warszawskiego Uniwersytetu Medycznego. http://www.wum.edu.pl/

Szmit, M. (2014). *Wybrane zagadnienia opiniowania sądowo-informatycznego. Wydanie 2, rozszerzone i poprawione*. Polskie Towarzystwo Informatyczne.

Szmit, M., Baworowski, A., Kmieciak, A., Krejza, P., Niemiec, A. (2011). *Elementy informatyki sądowej*. Polskie Towarzystwo Informatyczne.

Vitorino, P., Avila S., Perez, M., A. Rocha, A. (2018). Leveraging deep neural networks to fight child pornography in the age of social media. *Journal of Visual Communication and Image Representation*, *50*, 303–313. http://dx.doi.org/10.1016/j.jvcir.2017.12.005

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A. (2016). *Learning deep features for discriminative localization*. Proceedings of the IEEE International Conference on Computer Vision.

Description and usage of the FenrirAI applications for recognizing child sexual abuse materials

Child Sexual Abuse Material (CSAM) includes photographs, videos, drawings, computer-generated images in which the image of a person under the age of eighteen is captured or created during a sexual act and/or displaying nudity. Experts in the fields of computer science, anthropology, and sexology play a significant role in the assessment of such content. Due to the increasing number of initiated proceedings and cases related to accessing, disseminating, or creating CSAM, computer research is being conducted to develop tools that facilitate the evaluation of these materials. While solutions supporting the identification of content depicting child sexual abuse have been present for some time, recent research has particularly focused on the use of machine learning, known as 'artificial intelligence', to optimize the classification process. The aim of this article is to describe the series of publications comprising the process of creating the first Polish applications (FenrirAI) for recognizing content depicting child sexual abuse, describing their applications from a sexological and legal perspective, and presenting their limitations. Other studies related to the use of machine learning for automatic detection of pornographic materials will also be discussed, as well as future directions for software development. FenrirAI applications are provided free of charge to forensic experts, law enforcement officers, and researchers. The project documentation is available under the following DOI: 10.17605/OSF.IO/RU7JX.

KEYWORDS

CHILD SEXUAL ABUSE MATERIAL, CSAM, COMPUTER FORENSICS, MACHINE LEARNING

Cytowanie:

Oronowicz-Jaśkowiak, W. (2023). Opis i zastosowanie aplikacji FenrirAI do rozpoznawania treści przedstawiających seksualne wykorzystywanie dzieci *Dziecko Krzywdzone. Teoria, badania, praktyka, 22*(3), 114–130.

Fundusz Sprawiedliwości

Ministerstwo Sprawiedliwości

Sfinansowano ze środków Funduszu Sprawiedliwości, którego dysponentem jest Minister Sprawiedliwości