




EWA ALEKSANDRA PŁOCHA
Uniwersytet Kardynała Stefana Wyszyńskiego
w Warszawie

 <https://orcid.org/0000-0002-2245-4295>

Narzędzia oceny ryzyka powrotu do przestępstwa oparte na sztucznej inteligencji: uwagi na przykładzie aplikacji COMPAS

The artificial intelligence-based recidivism risk assessment instruments: Remarks on the example of the COMPAS application

Abstract: Over recent years, recidivism risk assessment tools have evolved and have been modified by using artificial intelligence technologies. With the aim of speeding up proceedings and introducing neutral mechanisms to guarantee equal treatment, these instruments are increasingly being implemented in criminal justice systems. They have a large number of both supporters and opponents. One of the best-known tools of this kind is COMPAS, which is used extensively in the United States of America for more than two decades. During this time, many scientific studies have been produced on it, but also a number of judgements have been issued that allow a reliable and objective assessment of the usefulness of algorithms of this kind and the identification of problems that may arise from their applying.

Keywords: artificial intelligence, risk assessment instruments, recidivism, criminological prognosis, COMPAS

1. Wprowadzenie

W ostatnich latach technologie sztucznej inteligencji (ang. *artificial intelligence*, AI) są coraz częściej wdrażane w systemach wymiaru sprawiedliwości w sprawach karnych na całym świecie. Wśród nich na znaczeniu zyskują algorytmy wspomagające podejmowanie decyzji przez człowieka, oparte na modelach predykcyjnych, które mogą być wykorzystywane

m.in. do przewidywania zachowań (w tym przestępstw i innych niepożądanych incydentów, takich jak samookaleczanie się skazanych, zaśląbnięcia, próby samobójcze, ucieczki czy bójki w zakładach karnych) lub profilowania potencjalnych pokrzywdzonych i sprawców.

Wymaga tu odnotowania, że algorytmy AI ze względu na swoją funkcję predykcyjną znalazły szerokie zastosowanie w narzędziach oceny ryzyka powrotu do przestępstwa (ang. *risk assessment tools*, RAT; *risk assessment instruments*, RAI), za pomocą których sporządza się m.in. prognozy kryminologiczne w odniesieniu zarówno do środków karnych, jak i do środków zabezpieczających. Związane jest to ze wzrostem znaczenia polityki kryminalnej opartej na dowodach naukowych (ang. *evidence-based crime policy*)¹ oraz propagującej stosowanie wystandaryzowanych narzędzi szacowania ryzyka recydywy² i potrzeb kryminogennych sprawcy w celu uzyskania dokładnych informacji wymaganych w przypadku wyrokowania, wykonania kary pozbawienia wolności czy „zwolnienia” z niej.

Wprawdzie obecnie narzędzia tego rodzaju zostały już implementowane w wielu państwach (m.in. program OASys w Estonii³ i w Wielkiej Brytanii, Sexual Violence Risk-20 i Risk for Sexual Violence Protocol w Australii), lecz na szczególną uwagę zasługuje aplikacja COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), wykorzystywana na dużą skalę w Stanach Zjednoczonych Ameryki, która należy nie tylko do najbardziej znanych, lecz także do najdłużej użytkowanych algorytmów predykcyjnych służących do oceny ryzyka recydywy. Pomimo że omawianą aplikację wdrożono w systemie *common law*, rekomendowane rozwiązania, ale też opinie, zastrzeżenia czy wątpliwości zgłaszane przez badaczy z zakresu różnych dziedzin nauki oraz praktyków pozostają uniwersalne i powinny być brane pod rozwagę zarówno przez państwa, które zamierzają implementować tego typu narzędzia, jak i przez państwa już stosujące podobne instrumenty, niezależnie od specyfiki danego porządku prawnego.

Trzeba nadmienić, że również w Polsce rozważane są możliwości wdrożenia narzędzi oceny ryzyka opartych na AI w wymiarze sprawied-

¹ Zakłada ona odstępianie od empirycznej prognozy indywidualnej (prognozy klinicznej) na rzecz zobiektywizowanej uzewnętrznionej oceny na podstawie danych statystycznych – zamiast wielu zob. K. Mamak: *Prognoza kryminologiczna*. W: Idem: *Rewolucja cyfrowa a prawo karne*. Kraków 2019, s. 101 i nn.

² Wprawdzie w poszczególnych systemach prawnych pojęcie recydywy jest definiowane w różnicowany sposób, lecz na potrzeby niniejszego opracowania przyjęto jego ogólne, uproszczone rozumienie jako „ponowienie jakiegokolwiek przestępstwa”.

³ Szerzej na ten temat – zob. K. Burdziak, E. Rützel: *Prognoza kryminologiczna. Porównanie rozwiązań estońskich i polskich*. Lublin 2022, s. 49–66.

liwości w sprawach karnych⁴. W naszym porządku prawnym szacowanie prawdopodobieństwa następuje w przypadkach wskazanych w ustawie na każdym etapie procesu karnego (w postępowaniu przygotowawczym, jurysdykcyjnym, w tym podczas wyrokowania, oraz wykonawczym)⁵.

⁴ Zob. m.in. D. Wójcik: *Stosowanie w postępowaniu karnym narzędzi diagnostyczno-prognostycznych służących oszacowaniu ryzyka powrotności do przestępstwa*. „Prawo w Działaniu. Sprawy karne” 2013, nr 16, s. 59–102; M. Kowalewska-Lukuć, K. Burdziak: *The use of AI in formulating a criminological prognosis of an offender*. „Technium Social Sciences Journal” 2021, vol. 25, s. 115–125; K. Burdziak, E. Rüütel: *Prognoza kryminologiczna...*, s. 67–88.

⁵ Elementy prognostyczne dotyczą w szczególności: stosowania środków zapobiegawczych (np. art. 249 § 1 ustawy z dnia 6 czerwca 1997 r. – Kodeks postępowania karnego. T.j. Dz.U. 2022, poz. 1375 ze zm. [dalej: k.p.k.]; art. 250 § 2a k.p.k.; art. 258 § 1 i 3 k.p.k.; art. 259 § 1 pkt 1 i 2 k.p.k.; art. 275a § 1 k.p.k.; art. 277 § 1 k.p.k.); wyrokowania w przedmiocie środków związanych z poddaniem sprawcy próbie (warunkowe umorzenie postępowania: art. 66 § 1 ustawy z dnia 6 czerwca 1997 r. – Kodeks karny. T.j. Dz.U. 2022, poz. 1138 ze zm. [dalej: k.k.]; warunkowe zawieszenie wykonania kary: art. 69 § 1 k.k.; warunkowe przedterminowe zwolnienie: art. 77 § 1 k.k.) i środków zabezpieczających (art. 93b § 1 k.k.; art. 93b § 3 k.k.; art. 93c pkt 4 k.k., art. 93g k.k.), a także wskazania typu zakładu karnego (art. 88a ustawy z dnia 6 czerwca 1997 r. – Kodeks karny wykonawczy. T.j. Dz.U. 2023, poz. 127 ze zm. [dalej: k.k.w.]); objęcia skazanego szczególną ochroną (art. 88d § 1 k.k.w.); kwalifikowania skazanych (sprawców) pozostających pod dozorem do grup ryzyka powrotu do przestępstwa (art. 169b k.k.w.); skierowania do zakładu psychiatrycznego o odpowiednich warunkach zabezpieczenia (art. 200a i art. 200b k.k.w.); zezwolenia na czasowy pobyt poza zakładem psychiatrycznym (art. 204d § 1 k.k.w.) czy wreszcie orzekania w przedmiocie uchylenia lub zmiany środka zabezpieczającego w postępowaniu karnym wykonawczym. W tym miejscu należy wyjaśnić, że w Polsce wymóg szacowania ryzyka powrotu do przestępstwa przez sprawcę oddanego pod dozór kuratora sądowego został wprowadzony na mocy przepisów § 9 ust. 3 pkt 3 i ust. 4, § 10 ust. 1 pkt 3 i ust. 2, § 11–14 rozporządzenia Ministra Sprawiedliwości z dnia 26 lutego 2013 r. w sprawie sposobu wykonywania obowiązków i uprawnień przez kuratorów sądowych w sprawach karnych wykonawczych (Dz.U. 2013, poz. 335), które weszło w życie 26 marca 2013 r. Szerzej – zob. B. Stańdo-Kawecka: *Wybrane problemy profesjonalizacji organów probacyjnych i klasyfikacji sprawców oddanych pod dozór do grup ryzyka*. „Nowa Kodyfikacja Prawa Karnego” 2014, t. 33, s. 11–39. Jednakże wyrokiem z dnia 16 grudnia 2014 r. (U 2/14. OTK-A 2014, nr 11, poz. 119) Trybunał Konstytucyjny uznał § 11–14 tegoż rozporządzenia za niezgodne z art. 92 ust. 1 zdanie pierwsze Konstytucji Rzeczypospolitej Polskiej. Wyjaśnił bowiem, że „niedopuszczalne jest zarówno wprowadzenie w przepisach rangi podstawowej zasad klasyfikacji skazanych oddanych pod dozór, jak i określenie obowiązków kuratora wynikających z przydzielenia skazanych do określonej grupy. Podział skazanych na określone grupy ryzyka powrotu do przestępstwa i określenie obowiązków kuratorów sądowych w zależności od przynależności skazanego do grupy stanowi jednocześnie określenie praw i obowiązków skazanych – obowiązków tym dolegliwszych, im surowsza kwalifikacja” (ibidem, teza 5.5). Konsekwencją tego orzeczenia było przeniesienie wymienionych zagadnień do Kodeksu karnego wykonawczego poprzez dodanie art. 169b (obowiązującego od 1 lipca 2015 r.) ustawą z dnia 20 lutego 2015 r. o zmianie ustawy – Kodeks karny oraz niektórych innych ustaw (Dz.U. 2015, poz. 396). Natomiast pozostałe kwestie reguluje obecnie rozporządzenie Ministra Sprawiedliwości z dnia 3 czerwca 2016 r.

Stosownie do kompetencji przewidzianych w ustawie decyzje procesowe podejmują odpowiednio m.in. prokurator bądź sąd, kierując się zasadą swobodnej oceny dowodów (art. 7 k.p.k.) – także jeśli chodzi o analizę prognostyczną. Organy postępowania karnego zatem kształtują swe przekonanie w tym obszarze na podstawie wszystkich przeprowadzonych dowodów (zwłaszcza opinii biegłych, o ile zostały zlecone – metoda kliniczna), ocenianych swobodnie z uwzględnieniem zasad prawidłowego rozumowania oraz wskazań wiedzy i doświadczenia życiowego (tzw. metoda quasi-kliniczna)⁶. Procedura ta wywołuje jednak uwagi krytyczne – podkreśla się m.in. brak szczegółowego uzasadnienia decyzji procesowych, zależność od nastawienia sędziego, niemożliwość weryfikacji sposobu oceny ani określenia ważności każdego z kryteriów, a także ustalenia, czy zbadano wszystkie czynniki wymienione przez ustawodawcę. Podnoszą się głosy, że należy wprowadzić bardziej obiektywne, ustandaryzowane mierniki, rzeczywiście realizujące zasadę prawa do sądu w rozumieniu art. 45 ust. 1 Konstytucji Rzeczypospolitej Polskiej⁷. Dlatego w przypadku zjawiska powrotności do przestępstwa konieczne jest przeanalizowanie, jakie problemy prawne mogą wynikać ze stosowania narzędzi oceny ryzyka recydywy opartych na AI.

Aby udzielić odpowiedzi na powyższe pytanie, w opracowaniu pokrótce przedstawiono historię narzędzi oceny ryzyka recydywy, ogólnie scharakteryzowano oprogramowanie COMPAS, a następnie zaprezentowano wybrane orzeczenia sądów amerykańskich dotyczące korzystania z tej aplikacji w sprawach karnych. Ponadto dokonano przeglądu zgłaszanych uwag krytycznych i zastrzeżeń oraz odniesiono się do nich.

2. Geneza narzędzi oceny ryzyka recydywy

Początki przewidywania ryzyka kryminogennego w sądownictwie karnym sięgają pierwszej połowy XX w. Wprawdzie nie istniały wtedy jeszcze narzędzia diagnostyczne, lecz prognozy kryminologiczne dotyczące warunkowego zwolnienia z kary pozbawienia wolności były sporządzane przez kuratorów, funkcjonariuszy Służby Więziennej, psychologów, psychiatrów, pracowników społecznych czy sędziów za pomocą metody klinicznej, która polegała na profesjonalnym osądzie (ang. *professional*

w sprawie sposobu i trybu wykonywania czynności przez kuratorów sądowych w sprawach karnych wykonawczych (Dz.U. 2016, poz. 969 ze zm.).

⁶ Szerzej na temat wskazanych metod prognozowania – zob. K. Mamak: *Prognoza kryminologiczna...*, s. 101 i nn.

⁷ Konstytucja Rzeczypospolitej Polskiej z dnia 2 kwietnia 1997 r. Dz.U. 1997, nr 78, poz. 483 ze sprst. i zm.

judgment) opartym na własnym doświadczeniu, intuicji i wyszkoleniu zawodowym⁸ (tzw. pierwsza generacja oceny ryzyka – ang. *first generation risk assessment*).

Z uwagi na naturalne błędy ludzkie i poznawcze poszukiwano bardziej miarodajnych i obiektywnych rozwiązań⁹. Próby opracowania narzędzi prognostycznych na potrzeby oceny ryzyka powrotu do przestępstwa podjęto najwcześniej w Stanach Zjednoczonych Ameryki¹⁰. Druga generacja szacowania ryzyka (ang. *second generation risk assessment*) miała swoje początki w latach 20. XX w. – w 1928 r. E.W. Burgess stworzył narzędzie prognostyczne, opierając się na metodzie aktuarialnej (statystycznej)¹¹, która polega na numerycznej predykcji dokonanej na podstawie analizy statycznych czynników ryzyka (ang. *static risk factors*). Pierwsze narzędzie aktuarialne powstało na potrzeby systemu zwolnień warunkowych w Illinois, a metoda aktuarialna do dzisiaj jest uznawana za najbardziej miarodajną i z tego względu nadal stosowana m.in. w amerykańskim wymiarze sprawiedliwości¹². Pomimo że omawiane instrumenty opracowano tam tak wcześnie, stały się one powszechne w organach wymiaru sprawiedliwości w sprawach karnych dopiero w latach 70. i 80.

Trzecia generacja narzędzi oceny ryzyka także opiera się na metodzie aktuarialnej, lecz analizie są poddawane zarówno statyczne, jak i dynamiczne czynniki ryzyka, które odgrywają istotną rolę szczególnie w przypadku przestępstw z użyciem przemocy i na tle seksualnym. Instrumenty te służą m.in. do mierzenia efektywności resocjalizacji¹³, a do owej grupy należą np. Assessment, Case Management and Evaluation System (ACE) czy Level of Service Inventory – Revised (LSI-R)¹⁴.

W ostatnich latach wykształciła się czwarta generacja narzędzi oceny ryzyka, łączących „systematyczną interwencję i monitorowanie z oceną szerszego zakresu czynników ryzyka, które wcześniej nie były mierzone, oraz innych czynników osobistych ważnych dla leczenia”¹⁵. Mają one

⁸ B. Stańdo-Kawecka: *Ruch „What Works” i „technologie szacowania ryzyka”*. W: Eadem: *Polityka karna i penitencjarna między punitywizmem i menedżeryzmem*. Warszawa 2020, s. 227–235.

⁹ *Public Safety Risk Assessment Clearinghouse: History of risk assessment*. Źródło: strona WWW Bureau of Justice Assistance, U.S. Department of Justice: <https://bja.ojp.gov/program/psrac/basics/history-risk-assessment#gov4e> [dostęp: 25.08.2023].

¹⁰ D. Wójcik: *Stosowanie w postępowaniu karnym narzędzi diagnostyczno-prognostycznych...*, s. 59.

¹¹ J. Rivees, cyt. za: K. Burdziak, E. Rüütel: *Prognoza kryminologiczna...*, s. 50.

¹² *Public Safety Risk Assessment Clearinghouse...*

¹³ Szerzej na ten temat – zob. D. Wójcik: *Stosowanie w postępowaniu karnym narzędzi diagnostyczno-prognostycznych...*, s. 64–68.

¹⁴ *Ibidem*, s. 70.

¹⁵ K. Burdziak, E. Rüütel: *Prognoza kryminologiczna...*, s. 51.

maksymalizować korzyści z terapii i nadzoru poprzez m.in. tworzenie planów wykonywania szeroko rozumianych sankcji karnych (zarządzanie sprawą – ang. *case management*), dostosowanych do indywidualnego profilu ryzyka i potrzeb skazanego (ang. *risk and needs profile*). Jednym z takich instrumentów jest COMPAS (Correctional Offender Management Profiling for Alternative Sanctions).

3. Charakterystyka aplikacji COMPAS

Oprogramowanie COMPAS wyprodukowano i wprowadzono na rynek w 1998 r.¹⁶ i od tamtej pory ciągle się je udoskonala (koryguje oraz optymalizuje), aby m.in. zapewnić większą przejrzystość, obiektywność i dokładność przewidywań. Okresowe aktualizacje służą wdrożeniu dobrych praktyk, dostosowaniu tego narzędzia do rozwoju nauki (zwłaszcza osiągnięć kryminologii i statystyki) i do nowych technologii (np. sztucznej inteligencji). Omawiana aplikacja jest jednym z algorytmów predykcyjnych najczęściej wykorzystywanych przez organy systemu wymiaru sprawiedliwości w sprawach karnych w USA (m.in. w stanach Nowy Jork, Wisconsin, Kalifornia i Floryda)¹⁷. Za pomocą programu COMPAS opracowuje się oceny ryzyka recydywy i potrzeb, które należą do czynników branych pod uwagę przy podejmowaniu decyzji procesowych w postępowaniu przygotowawczym (odnośnie do stosowania środków zapobiegawczych izolacyjnych i nieizolacyjnych), jurysdykcyjnym (szczególnie w fazie wyrokowania – wybór pomiędzy sankcjami o charakterze wolnościowym a karą pozbawienia wolności, orzekanie co do środków związanych z poddaniem sprawcy próbie i środków zabezpieczających) i wykonawczym (np. rozstrzygnięcie w przedmiocie warunkowego przedterminowego zwolnienia). Pozwala on na efektywne zarządzanie sprawą i stworzenie indywidualnego profilu ryzyka i potrzeb danej osoby. Wygenerowana w ten sposób ocena jest przydatna do ustalenia, czy przestępca może być bezpiecznie i skutecznie nadzorowany w społeczeństwie, a także do określenia warunków probacji, nadzoru i reakcji na

¹⁶ Informacja na oficjalnej stronie producenta: <https://www.equivant.com/faq/> [dostęp: 31.08.2023]. Zostało wyprodukowane przez spółkę Northpoint Inc. (obecnie funkcjonującą pod nazwą Equivant i stanowiącą własność Constellation Software Inc.), którą założyli w 1989 r. Tim Brennan, profesor statystyki na Uniwersytecie Kolorado, i Dave Wells, prowadzący program korekcyjny w Traverse City w stanie Michigan.

¹⁷ A. „Mac” Taylor: *AI prediction tools claim to alleviate an overcrowded American justice system... but should they be used?*. „Stanford Politics” [online], 13.09.2020. <https://stanfordpolitics.org/2020/09/13/ai-prediction-tools-claim-to-alleviate-an-overcrowded-american-justice-system-but-should-they-be-used/> [dostęp: 25.08.2023].

naruszenia (np. ustalenia częstotliwości składania raportów, badania na obecność środków odurzających, opracowania strategii leczenia, zastosowania monitoringu elektronicznego czy zezwolenia na prace społeczne)¹⁸. Oprócz tego aplikacja może być używana do „poprawy oceny, ważenia i stosowania przez sędziego innych dowodów skazujących w formułowaniu zindywidualizowanego programu karnego odpowiedniego dla każdego oskarżonego”¹⁹.

Producent opisuje oprogramowanie COMPAS jako zestaw dostępnych przez Internet narzędzi oceny ryzyka i potrzeb (pod nazwą Northpointe Suite), zaprojektowanych do wspierania organów systemu wymiaru sprawiedliwości w sprawach karnych w podejmowaniu decyzji w przedmiocie umieszczenia przestępców w zakładach karnych, nadzoru nad nimi i zarządzania ich sprawami²⁰. COMPAS składa się z modułu podstawowego, zwanego COMPAS Core, a także modułów dodatkowych przeznaczonych do postępowania przygotowawczego, sądowego i wykonawczego (np. Recidivism Risk Screen, Pretrial Release Risk Scale II)²¹, które są wdrażane na zlecenie zamawiającego. Działanie aplikacji opiera się na wyodrębnionych skalach ryzyka i skalach potrzeb. Pierwsze koncentrują się przede wszystkim na czynnikach związanych z recydywą (ryzyko recydywy ogólnej – ang. *general recidivism risk*; ryzyko recydywy z użyciem przemocy – ang. *violent recidivism risk*), drugie zaś – na zakresie i rodzajach interwencji terapeutycznych, które byłyby najkorzystniejsze dla konkretnego przestępcy.

Dane na temat sprawcy są uzyskiwane z akt sprawy karnej oraz w toku ustrukturyzowanego wywiadu przeprowadzanego z nim za pomocą kwestionariusza (COMPAS Core) złożonego ze 137 pytań²², które dotyczą zarówno faktów, jak i ocen. Na podstawie tych informacji generowany jest raport²³ składający się z dwóch części:

¹⁸ B. Stańdo-Kawecka: *Korzystanie z informacji o wynikach szacowania ryzyka i potrzeb przez sąd wyrokujący*. W: Eadem: *Polityka karna i penitencjarna...*, s. 244–245; *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), § 88–91.

¹⁹ *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), § 92 [tłum. własne].

²⁰ Zob. Equivant: *What is COMPAS?*. <https://www.equivalent.com/faq/> [dostęp: 25.08.2023].

²¹ Equivant: *Practitioner’s guide to COMPAS Core*. 4.04.2019, s. 12. <https://www.equivalent.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf> [dostęp: 2.09.2023].

²² Ze wzorem kwestionariusza stosowanym w stanie Nowy Jork można się zapoznać w dokumencie *New York State COMPAS – probation risk and need assessment study* na stronie WWW Division of Criminal Justice Services: https://www.criminaljustice.ny.gov/crimnet/ojsa/opca/compas_probation_report_2012.pdf [dostęp: 2.09.2023].

²³ Przykładowy raport jest dostępny na stronie WWW <https://assets.documentcloud.org/documents/2839240/Sample-Risk-Assessment-COMPAS-Results.pdf> [dostęp: 1.09.2023].

1. Ogólny potencjał ryzyka (ang. *overall risk potential*) – zawiera prognozę m.in. co do użycia przemocy, ponowienia przestępstwa, niestawienia. Wyniki z tej części mają charakter predykcyjny. Jak wskazują twórcy aplikacji, celem owej skali jest odróżnienie osób, które ponowią przestępstwo, od tych, które tego nie uczynią.
2. Profil kryminogeny i potrzeb (ang. *criminogenic and needs profile*) – obejmuje czynniki związane z historią kryminalną (ang. *criminal history factors*); ocenę potrzeb (ang. *needs assessment*): nadużywanie środków odurzających, problemy finansowe / ubóstwo, problemy zawodowe i edukacyjne; stosunek do przestępstwa (ang. *criminal attitude*); otoczenie społeczne (ang. *social environment*): przestępczość w rodzinie, otoczenie społeczne, czas wolny od pracy i wypoczynek, niestabilność mieszkaniowa, problemy z przystosowaniem społecznym; czynniki wyższego rzędu (ang. *higher order factors*): niepowodzenie w zakresie socjalizacji, okazja do popełnienia przestępstwa, osobowość przestępcza, izolacja społeczna. Wprawdzie ustalenia zawarte w tej części nie mają charakteru predykcyjnego, lecz przyczyniają się do określenia całkowitego ryzyka recydywy, gdyż pozwalają szczegółowo opisać sprawcę²⁴.

Wyniki obu części raportu mają postać wykresów słupkowych, a każdy słupek wskazuje poziom ryzyka w skali od 1 do 10 (od 1 do 4 – niskie, powyżej 4 do 7 – średnie, powyżej 7 do 10 – wysokie).

Trzeba dodać, że raport COMPAS zawiera prognozę wygenerowaną dzięki porównaniu uzyskanych informacji o sprawcy ze zbliżoną grupą danych, zatem algorytm nie przewiduje zindywidualizowanego, konkretnego prawdopodobieństwa ponowienia przestępstwa, a jedynie dokonuje ogólnej predykcji. W uproszczeniu można to przedstawić w ten sposób, że po wprowadzeniu informacji o przestępcy do aplikacji wyszukuje ona wszystkie osoby z takim samym zestawem cech jak u niego i przypisuje go do jednej z ośmiu kategorii sprawców (w klasyfikacji dla kobiet lub dla mężczyzn)²⁵, a następnie na podstawie posiadanych statystyk na temat osób „podobnych” (np. ile z nich zostało aresztowanych w okresie warunkowego zawieszenia wykonania kary, a ile nie) szacuje ryzyko ponowienia przestępstwa²⁶.

²⁴ K. Schwerzmann: *Abolish! Against the use of risk assessment algorithms at sentencing in the US criminal justice system*. „Philosophy & Technology” 2021, no. 34, s. 1890. <https://doi.org/10.1007/s13347-021-00491-2>.

²⁵ Szczegółowe opisy dotyczące poszczególnych kategorii przestępców płci męskiej i żeńskiej – zob. Equivalent: *Practitioner’s guide...*, s. 54–61.

²⁶ Obszerne wyjaśnienia w tym zakresie zawiera np. L. Eckhouse et al.: *Layers of bias: A unified approach for understanding problems with risk assessment*. „Criminal Justice and Behavior” 2019, vol. 46, issue 2, s. 185–209. <https://doi.org/10.1177/0093854818811379>.

4. Wybrane orzeczenia sądów amerykańskich dotyczące aplikacji COMPAS

W Stanach Zjednoczonych Ameryki dopuszczalność użycia narzędzi oceny ryzyka powrotu do przestępstwa przez organy systemu wymiaru sprawiedliwości w sprawach karnych została uregulowana przede wszystkim w przepisach prawa poszczególnych stanów: w niektórych stanach korzystanie z takich narzędzi w określonych przypadkach jest obligatoryjne, a w innych – fakultatywne²⁷. Początkowo instytucje państwowe były nastawione sceptycznie do tych instrumentów i zdarzało się, że ich nie stosowały. Przykładowo w stanie Nowy Jork obowiązek brania pod uwagę wyników resocjalizacji i oceniania ryzyka za pomocą narzędzia COMPAS przez stanową komisję ds. zwolnień warunkowych wprowadzono na mocy przepisów ustawy o budżecie wykonawczym na lata 2011–2012²⁸, lecz wymagał on potwierdzenia w orzecznictwie sądowym. W sprawie *Diaz vs. New York State Board of Parole* wyjaśniono bowiem, że wspomniana komisja nadal musi przeprowadzać rzeczywistą, pogłębianą ocenę każdego wnioskodawcy, ale teraz ocena ta ma obejmować instrument, który mierzy resocjalizację i prawdopodobieństwo powodzenia zwolnienia warunkowego²⁹. Z kolei w sprawie *Morris vs. New York State Department of Corrections and Community Supervision et al.* za naruszenie prawa uznano brak rozważenia wewnętrznej oceny ryzyka (korzystnej dla osadzonego), mimo że została ona sporządzona i załączona do akt sprawy³⁰.

Należy nadmienić, że COMPAS – podobnie jak inne narzędzia tego rodzaju – ma jedynie charakter pomocniczy, a wygenerowana dzięki

²⁷ Przykładowo kilka stanów (Kentucky, Ohio, Pensylwania, Oklahoma) uchwaliło przepisy, które nakładają na właściwe organy obowiązek przekazania sędziom ocen ryzyka i danych w zakresie ponowienia przestępstwa na potrzeby wyrokowania, inne stany zaś zezwoliły na to, ale tego nie nakazały (np. Idaho, niektóre sądy w Luizjanie) – zob. *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), § 42. Na mocy First Step Act (FSA), uchwalonej w 2018 r., w federalnych zakładach karnych korzysta się z aplikacji PATTERN (ang. Prisoner Assessment Tool Targeting Estimated Risk and Needs) – zob. U.S. Department of Justice. Office of the Attorney General: *The First Step Act of 2018: Risk and needs assessment system*. <https://www.ojp.gov/First-Step-Act-of-2018-Risk-and-Needs-Assessment-System> [dostęp: 2.09.2023].

²⁸ Szerzej na ten temat – zob. J. Caher: *Effect of risk assessment rule on parole decisions is unclear*. Źródło: <https://www.cherylkatesesq.com/news/2018/9/6/effect-of-risk-assessment-rule-on-parole-decisions-is-unclear-april-30-2012-by-john-caher> [dostęp: 30.08.2023].

²⁹ *Diaz v. New York State Board of Parole*. Źródło: <https://law.justia.com/cases/new-york/other-courts/2013/2013-ny-slip-op-23422.html> [dostęp: 2.09.2023].

³⁰ *Morris v. New York State Department of Corrections and Community Supervision et al.* Źródło: <https://law.justia.com/cases/new-york/other-courts/2013/2013-ny-slip-op-50604-u.html> [dostęp: 2.09.2023].

niemu ocena ryzyka ponowienia przestępstwa nie jest wiążąca dla organów podejmujących decyzje procesowe: zawiera tylko wskazówki, które organ procesowy może uwzględnić (w całości lub w części) albo nie. Przykładowo w sprawie *Partee vs. Evans* stwierdzono, że choć sprawca uzyskał niski wynik według aplikacji COMPAS, dopuszczalne jest przypisanie większego znaczenia historii przemocy (a także wcześniejszemu uwięzieniu i niedawnemu naruszeniu poziomu III) oraz wydanie na tej podstawie decyzji odmownej, ponieważ w obowiązujących przepisach nie ma obowiązku omawiania każdego rozważanego czynnika i nadania każdemu równej wagi³¹. Zbliżone orzeczenie wydano w sprawie *Williams vs. New York*³². Otóż uznano, że można położyć większy nacisk na powagę przestępstwa i odmówić zwolnienia warunkowego pomimo niskiego stopnia prawdopodobieństwa w raporcie COMPAS, pod warunkiem iż wynik ten został wzięty pod uwagę.

Na temat dopuszczalności powołania się na ocenę ryzyka wygenerowaną za pomocą aplikacji COMPAS przy wydawaniu wyroku wypowiedziano się zdawkowo w sprawie *State vs. Samsa*. Wprawdzie kwestia ta nie była analizowana przez pryzmat ewentualnego naruszenia prawa do sądu, lecz stwierdzono, że „COMPAS jest tylko jednym z narzędzi dostępnych dla sądu w czasie orzekania”³³. Uszczegółowienie tego zagadnienia zawarto w orzeczeniu dotyczącym aplikacji LSI-R w sprawie *Malenchik vs. State*, w którym wyjaśniono, że oceny ryzyka recydywy generowane przez oprogramowanie tego rodzaju: 1) nie mogą stanowić okoliczności obciążającej ani łagodzącej, 2) nie mogą służyć do wyznaczenia całkowitego okresu trwania kary, ale mogą być używane do ustalenia sposobu jej odbycia³⁴. Zastrzeżenia te mają charakter fundamentalny i zostały potwierdzone w kolejnych orzeczeniach sądowych.

W odniesieniu do aplikacji COMPAS najczęściej przywołuje się orzeczenie wydane w sprawie *State vs. Loomis*³⁵, w którym stwierdzono – najogólniej rzecz ujmując – że korzystanie z tego rodzaju narzędzia przy wydawaniu wyroków, z zastrzeżeniem pewnych ograniczeń, nie narusza prawa oskarżonego do rzetelnego procesu sądowego, chociaż metodologia zastosowana do sporządzenia oceny nie została ujawniona ani sądowi, ani oskarżonemu. Trzeba wyjaśnić, że skarżący podniósł zarzut naruszenia

³¹ *Partee v. Evans*, 40 Misc.3d 896.

³² *Williams v. New York*, 49 Misc.3d 732.

³³ *State v. Samsa*, 2015 WI App 6, 359 Wis.2d580, 859 N.W.2d149, § 13 [tłum. własne].

³⁴ *Malenchik v. State*, 928 N.E.2d 564, 574 (Ind. 2010).

³⁵ *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016). Zob. *Criminal law – Sentencing guidelines – Wisconsin Supreme Court requires warning before use of algorithmic risk assessments in sentencing – State v. Loomis*, 881 N.W.2d 749 (Wis. 2016). „Harvard Law Review” 2017, vol. 130, issue 5, s. 1530–1537.

jego prawa do rzetelnego procesu sądowego poprzez samo posłużenie się podczas wyrokowania w jego sprawie oceną ryzyka wygenerowaną przez COMPAS. Wskazał trzy okoliczności. Po pierwsze, ponieważ w aplikacji zastosowano metodologię, która jest objęta tajemnicą handlową i nie została ujawniona w toku postępowania, nie można było zweryfikować naukowej ważności oceny ryzyka (nie wiadomo, jak są ważone czynniki, przypisywane punkty i określone wyniki), a tym samym – dokładności raportu, wskutek czego naruszono prawo skarżącego do bycia skazanym na podstawie dokładnych informacji; po drugie, naruszono jego prawo do zindywidualizowanego wyroku (przez odniesienie się do danych grupowych); po trzecie, zaskarżone orzeczenie naruszyło prawo oskarżonego do niebycia skazanym na podstawie płci, gdyż ocena ryzyka recydywy jest generowana z uwzględnieniem tej cechy.

Kluczową okolicznością, która zadecydowała o oddaleniu wymienionych zarzutów, było stwierdzenie sądu okręgowego, że tożsamy wyrok zapadłby nawet bez oceny ryzyka z aplikacji COMPAS. Niezależnie od tego samo podniesienie zarzutu naruszenia prawa do rzetelnego procesu sądowego dało Sądowi Najwyższemu Wisconsin możliwość pogłębionej analizy owego zagadnienia. Jednakże argumentacja przedstawiona w uzasadnieniu orzeczenia spotkała się z poważną krytyką. Za nieprzekonujące uznano bowiem twierdzenie sądu, że choć skarżący nie mógł przeanalizować, jak algorytm oblicza ryzyko, to do naruszenia jego praw nie doszło, ponieważ udostępniono mu treść raportu i mógł zakwestionować wyniki oceny zaprezentowane w postaci słupków. Trzeba zaznaczyć, że oprócz wykresu słupkowego raport COMPAS nie zawiera uzasadnienia ani wytłumaczenia, w jaki sposób oszacowano wartość poszczególnych wskaźników i dlaczego. W odniesieniu do tych kwestii Sąd Najwyższy Wisconsin odesłał skarżącego do *Przewodnika dla praktyków po aplikacji COMPAS*, w którym wbrew pozorom znajdują się tylko powierzchowne informacje. Dodatkowo organ nadmienił, że niektóre stany korzystające z tego algorytmu przeprowadziły jego badania walidacyjne i nie wykazały one nieprawidłowości. Jednakże należy wyjaśnić, iż każdy ze stanów dysponuje odrębnymi zestawami danych do „trenowania” algorytmu, uwzględniającymi specyfikę miejscowych populacji i subpopulacji. Ponadto Sąd Najwyższy Wisconsin zakwestionował badania naukowe, na które powołał się skarżący.

Co istotne, w omawianym orzeczeniu wskazano, że raport PSI (ang. *pre-sentence investigation*) z oceną ryzyka wygenerowaną przez COMPAS musi zawierać ostrzeżenie dla sądów dotyczące jej dokładności i obejmujące m.in. następujące informacje:

1. Szczegóły na temat działania aplikacji są objęte tajemnicą handlową, w tym co do sposobu ważenia czynników lub określania punktacji ryzyka.

2. Ocena ryzyka porównuje oskarżonych z próbą krajową, ale nie zakończono jeszcze badania walidacji krzyżowej dla populacji Wisconsin.
3. Niektóre badania wyników oceny ryzyka COMPAS wzbudziły wątpliwości co do tego, czy nie klasyfikują nieproporcjonalnie przestępców z grup mniejszościowych jako bardziej narażonych na recydywę.
4. Narzędzia oceny ryzyka muszą być stale monitorowane i ponownie normalizowane pod względem dokładności, gdyż populacje i subpopulacje się zmieniają³⁶.

Zastrzeżenia te powinno się regularnie aktualizować.

Co więcej, w analizowanym orzeczeniu – w kontekście zarzutu naruszenia prawa do zindywidualizowanego wyroku – stwierdzono, że gdyby ocena ryzyka COMPAS stanowiła decydujący czynnik brany pod uwagę podczas wyrokowania, to zarzut skarżącego mógłby być zasadny, ponieważ ów algorytm został zaprojektowany do ewaluacji danych grupowych (poniekąd przypomina aktuarialną ocenę ryzyka ubezpieczeniowego). Innymi słowy: ocena ryzyka COMPAS polega na przewidywaniu zachowań grupowych, a nie indywidualnych.

Ponadto według Sądu Najwyższego Wisconsin „uwzględnienie oceny ryzyka COMPAS wraz z innymi czynnikami pomocniczymi przy wydawaniu wyroku jest przydatne w dostarczeniu sądowi skazującemu jak największej ilości informacji, aby móc ustalić zindywidualizowany wyrok”³⁷, pozwala bowiem na ocenę wszystkich czynników, które zgodnie z ustawą należy brać pod uwagę. Wreszcie Sąd Najwyższy Wisconsin uznał, że wbrew zarzutom skarżącego płęć nie jest wykorzystywana jako czynnik kryminogeny, lecz służy wyłącznie do normowania statystycznego, i powołał się tu na wyniki badań naukowych. Jego zdaniem skarżący nie przedstawił wystarczających dowodów na to, że sąd skazujący faktycznie wziął pod uwagę płęć i doszło w tym zakresie do dyskryminacji.

O ile początkowo organy wymiaru sprawiedliwości w sprawach karnych niechętnie stosowały aplikację COMPAS, o tyle wraz z upływem czasu tendencja ta uległa odwróceniu, chociaż zastrzeżenia i wątpliwości dotyczące korzystania z oprogramowania pozostały. Trzeba zaznaczyć, że na podstawie przepisów karnoprosesowych wielokrotnie występowano do sądu o zobowiązanie producenta, aby przedłożył do akt jego szczegółową dokumentację techniczną. Jednakże w przeważającej mierze wnioski tego rodzaju nie były uwzględniane, w związku z czym próbowano uzyskać dokumentację techniczną narzędzia COMPAS w ramach dostępu do informacji publicznej. Przykładowo w sprawie *Rayner vs. State* zainicjowano postępowanie m.in. na mocy przepisów ustawy o wolności

³⁶ Ibidem, § 66.

³⁷ Ibidem, § 72 [tłum. własne].

informacji (ang. Freedom of Information Law, FOIL). Martha Rayner, profesor prawa z Fordham Law School³⁸, zwróciła się do Departamentu Więziennictwa i Nadzoru Społecznego Stanu Nowy Jork (ang. New York State Department of Corrections and Community Supervision, DOCCS) o udostępnienie wszelkiej dokumentacji na temat działania i solidności modułu COMPAS Re-entry, który jest obligatoryjnie stosowany przez Komisję ds. Zwolnień Warunkowych Stanu Nowy Jork w każdym postępowaniu o udzielenie zwolnienia warunkowego. DOCCS odrzucił jednak wnioski Rayner „w celu ochrony praw autorskich i interesów biznesowych prywatnej, nastawionej na zysk korporacji, która opracowała narzędzie COMPAS – Northpointe Inc.”³⁹. Skarżąca podniosła również, że Komisja ds. Zwolnień Warunkowych niemal zawsze przychyliła się do przewidywań COMPAS Re-entry dotyczących wysokiego ryzyka. W uzasadnieniu skargi wskazano, że opinia publiczna nie wie:

- 1) w jaki sposób algorytm ocenia i łączy odpowiedzi na pytania z kwestionariusza w celu wygenerowania oceny ryzyka i potrzeb;
- 2) jaki wynik jest wymagany, by przypisano go do danego stopnia w skali od 1 do 10;
- 3) w jaki sposób i dlaczego przyjęto przedziały dla poziomów ryzyka (niskie, średnie i wysokie);
- 4) jaka jest grupa danych podstawowych tworzących normę stosowaną do generowania wyników związanych z każdym decylem.

Skarżąca zawnioskowała m.in. o zobowiązanie Departamentu Więziennictwa i Nadzoru Społecznego Stanu Nowy Jork do przedstawienia żądanej dokumentacji w terminie 10 dni roboczych.

Po przesłuchaniu przedstawiciela zespołu badawczego producenta oprogramowania Sąd Najwyższy Hrabstwa Albany (ang. Albany County Supreme Court) uznał, że żądana dokumentacja techniczna zawiera informacje stanowiące tajemnicę handlową (ang. *trade secret*)⁴⁰. Według sądu opracowanie formuł, algorytmów, skal ryzyka i potrzeb oraz danych normalizacyjnych, na których opiera się narzędzie COMPAS, a także sposobu, w jaki wyniki są skorelowane z danymi z kwestionariusza Komisji

³⁸ Skarżąca współkieruje studencką poradnią prawną Criminal Defense Clinic, która głównie reprezentuje oskarżonych i skazanych w sprawach karnych.

³⁹ Treść złożonej skargi: https://law.yale.edu/sites/default/files/documents/pdf/mfia_rayner_nys.pdf [dostęp: 2.09.2023]. Zob. też M. Rayner: *Commentary: This algorithm is used to deny inmates parole. We're not allowed to know anything about it.* „Times Union” [online], 5.05.2023. <https://www.timesunion.com/opinion/article/commentary-proprietary-algorithm-deny-inmate-18001304.php> [dostęp: 2.09.2023].

⁴⁰ Rayner v. New York State Department of Corrections and Community Supervision. Źródło: <https://law.justia.com/cases/new-york/other-courts/2023/2023-ny-slip-op-23293.html> [dostęp: 23.01.2024].

ds. Zwolnień Warunkowych Stanu Nowy Jork, zidentyfikowanie wartości granicznych decyli i określenie metody przypisywania decyli do kategorii jakościowych wymagało poczynienia dużych inwestycji (czasu, środków pieniężnych, wysiłku) przez firmę Equivant – podmiot prywatny nastawiony na zysk. Zważywszy, że ujawnienie tej technologii bezsprzecznie ułatwiłoby potencjalnym konkurentom dostarczenie podobnego oprogramowania po znacznie niższych kosztach, należało stwierdzić, iż utrzymywanie jej w tajemnicy ma rzeczywistą wartość dla Equivantu i daje mu wyraźną przewagę nad konkurencją w stanie Nowy Jork (i być może w innych stanach). W rezultacie sąd uznał, że przedmiotowa technologia stanowi tajemnicę handlową w dobrej wierze (ang. *bona fide trade secret*), a dokumentacji technicznej, która ją opisuje i która została przekazana Departamentowi Więziennictwa i Nadzoru Społecznego Stanu Nowy Jork, nie obejmuje obowiązek udostępnienia (ujawnienia). Co więcej, sąd wyjaśnił, że aby skutecznie powoływać się na „wyjątkowo ważny interes publiczny” w postaci zapewnienia nadzoru nad zwolnieniem warunkowym jako na przesłankę uzyskania dostępu do informacji publicznej, wymagana jest zmiana prawa przez ustawodawcę stanowego⁴¹.

5. Uwagi krytyczne dotyczące aplikacji COMPAS

Pierwsze uwagi krytyczne, zastrzeżenia czy wątpliwości dotyczące oprogramowania COMPAS były zgłaszane niedługo po wdrożeniu go do amerykańskiego wymiaru sprawiedliwości w sprawach karnych. Pozwoliło to na wprowadzenie poprawek, aktualizację i rozwój aplikacji. Dodatkowo przyjęto model współdziałania oparty na partnerstwie publiczno-prywatnym, który umożliwia bieżące monitorowanie funkcjonowania tego narzędzia, przedstawianie uwag przez jego użytkowników, praktyków i badaczy, a także korygowanie i optymalizację oprogramowania w sposób ciągły⁴². Pomimo że wyniki badań i ewaluacji są udostępniane w postaci publikacji pokonferencyjnych, raportów technicznych, recenzowanych artykułów i rozdziałów w monografiach naukowych, zastosowanie aplikacji COMPAS w wymiarze sprawiedliwości w sprawach karnych nadal budzi szereg wątpliwości.

W 2016 r. organizacja pozarządowa ProPublica zaprezentowała raport, w którym stwierdzono, że COMPAS dyskryminuje osoby czarnoskóre: były one narażone na zawyżenie poziomu ryzyka recydywy

⁴¹ Ibidem.

⁴² Equivant: *Practitioner's guide...*, s. 12.

znacznie bardziej niż biali oskarżeni, tym drugim zaś program częściej go zaniżał⁴³. Autorzy zweryfikowali wszystkie oceny ryzyka ponowienia przestępstwa i potrzeb wygenerowane przez COMPAS w latach 2013 i 2014 w hrabstwie Broward na Florydzie (sporządzone łącznie dla 18 610 osób) oraz szczegółowo wytłumaczyli, w jaki sposób badali to narzędzie i jak doszli do niniejszych wniosków⁴⁴. Twórcy aplikacji COMPAS gruntownie przeanalizowali raport organizacji ProPublica i wykryli w nim błędy statystyczne i techniczne, które kolejno wyjaśnili, a następnie na tej podstawie zaprzeczyli twierdzeniom o dyskryminującym charakterze aplikacji w odniesieniu do osób czarnoskórych⁴⁵. Do podobnych wniosków doszli też inni badacze⁴⁶. W literaturze przedmiotu wskazuje się jednak, że w badaniach posłużono się różnymi metodologiami – stąd poczyniono odmienne ustalenia⁴⁷. Niemniej kwestia, czy dochodzi do dyskryminacji rasowej, nadal nie została definitywnie rozstrzygnięta.

W 2018 r. opublikowano wyniki kolejnego badania, którego uczestnicy (400 ochotników) niebędący ekspertami przewidywali ryzyko recydywy na podstawie informacji obejmujących zaledwie 7 danych o przestępcy i uzyskali wskaźnik dokładności (63%, po połączeniu odpowiedzi – 67%) zbliżony do poziomu osiągniętego przez aplikację COMPAS (65%), analizującą aż 137 cech/odpowiedzi sprawcy⁴⁸. Autorzy badania stworzyli nawet prosty algorytm, który działał równie dobrze jak COMPAS. Stoją oni na stanowisku, że amerykański wymiar sprawiedliwości gromadzi zbyt wiele informacji o sprawcach przestępstw. Ponadto zasygnalizowali wątpliwości, czy przewidywania algorytmów są faktycznie lepsze od ludzkich. Analogiczne konkluzje zostały sformułowane przez inne zespoły uczonych, jednak producenci oprogramowania COMPAS uznali, że opinie badaczy nie stanowią krytyki tego narzędzia, a w rzeczywistości potwier-

⁴³ J. Angwin et al.: *Machine bias. There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica, 23.05.2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [dostęp: 2.09.2023].

⁴⁴ J. Larson et al.: *How we analyzed the COMPAS recidivism algorithm*. ProPublica, 23.05.2016. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> [dostęp: 2.09.2023].

⁴⁵ Zob. T. Brennan, W. Dieterich, Ch. Mendoza: *COMPAS risk scales: Demonstrating accuracy equity and predictive parity*. Performance of the COMPAS risk scales in Broward County, 8.07.2016. Źródło: https://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf [dostęp: 25.08.2023].

⁴⁶ A.W. Flores, K. Bechtel, C. Lowenkamp: *False positives, false negatives, and false analyses: A rejoinder to „Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks”*. „Federal Probation” 2016, vol. 80, no. 2, s. 38–46.

⁴⁷ L. Eckhouse et al.: *Layers of bias: A unified approach...*, s. 185–209.

⁴⁸ Zob. J. Dressel, H. Farid: *The accuracy, fairness, and limits of predicting recidivism*, „Science Advances” 2018, vol. 4, issue 1, s. 1–5. <https://doi.org/10.1126/sciadv.aao5580>.

dzają, że osiąga ono dobrą przewidywalność i pasuje do coraz bardziej akceptowanego standardu AUC (0,70)⁴⁹.

W literaturze przedmiotu cały czas trwa dyskusja dotycząca stosowania aplikacji COMPAS i innych tego typu algorytmów predykcyjnych w wymiarze sprawiedliwości w sprawach karnych. Toczy się ona na wielu płaszczyznach i z udziałem przedstawicieli różnych dziedzin nauki. Złożoność materii wynika bowiem nie tylko ze specyfiki narzędzi szacowania ryzyka recydywy i użycia w nich technologii AI – wpływ na nią mają także obszar, w którym są wdrażane, i skutki, jakie niosą za sobą dla jednostek.

Należy wyjaśnić, że aby można było zapewnić poszanowanie konstytucyjnego standardu ochrony praw i wolności, narzędzia oceny ryzyka recydywy muszą być – podobnie jak sama AI – wiarygodne, czyli zgodne z prawem, etyczne i solidne (w ujęciu technicznym i społecznym)⁵⁰. Kwestia ta okazuje się jednak znacznie bardziej złożona. O ile bowiem metodę aktuarialną powszechnie uznaje się w środowisku naukowym za miarodajną, o tyle dobór danych, które podlegają analizie za jej pomocą w celu szacowania ryzyka ponowienia przestępstwa i potrzeb sprawy, budzi spory między kryminologami. Narzędzia drugiej generacji badały wyłącznie czynniki statyczne, na podstawie których opracowywano prognozę kryminologiczną. Kolejne generacje analizują również czynniki dynamiczne, co jest kwestionowane w literaturze przedmiotu. Podnosi się, że np. informacje te wykraczają poza zakres ustawy określony w prawie karnym materialnym oraz prowadzą pośrednio do dyskryminacji ze względu na płeć, rasę, przynależność do mniejszości etnicznej czy klasy społecznej (informacje o miejscu zamieszkania, wysokości zarobków, wykształceniu itp.)⁵¹. Trzeba zaznaczyć, iż nie ma tutaj zgodności pomiędzy samymi uczonymi, co już wystarcza, aby podważyć wiarygodność narzędzi szacowania ryzyka recydywy. Oprócz tego zastrzeżenia budzi ich dokładność. Wymagane jest ich ciągłe monitorowanie i uaktualnianie danych, zwłaszcza o zmieniających się populacjach i subpopulacjach.

Jednocześnie należy mieć na uwadze, że oparcie omawianych narzędzi na technologii sztucznej inteligencji skutkuje koniecznością spełnienia wymogów jej stawianych, czyli zapewnienia przewodniej i nadzorczej roli

⁴⁹ AUC – pole pod krzywą (ang. *area under curve*). E. Yong: *A popular algorithm is no better at predicting crimes than random people*. „The Atlantic” [online], 17.01.2018. <https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/> [dostęp: 2.09.2023].

⁵⁰ E.A. Płocha: *O pojęciu sztucznej inteligencji i możliwościach jej zastosowania w postępowaniu cywilnym*. „Prawo w Działaniu” 2020, nr 44, s. 280–284.

⁵¹ Zamiast wielu zob. L. Eckhouse et al.: *Layers of bias: A unified approach...*, s. 185–209.

człowieka, technicznej solidności i bezpieczeństwa, ochrony prywatności zarządzania danymi, przejrzystości (transparentności), różnorodności, niedyskryminacji i sprawiedliwości oraz odpowiedzialności⁵². W tym zakresie podnosi się – co zresztą ujawniają przywołane raporty z 2016 i 2018 r. – że w zasadzie nie wiadomo, jak działają aplikacja COMPAS i podobne do niej narzędzia. Brak jest informacji o tym, w jaki sposób:

- 1) algorytm porównuje dane sprawcy z określoną populacją,
- 2) przygotowywane są dane treningowe (problem tzw. uprzedzeń w danych),
- 3) tworzy się grupy,
- 4) przypisuje się wagi poszczególnym cechom (kryteriom).

Dostęp do informacji na ten temat dotychczas był niemożliwy z uwagi na objęcie ich tajemnicą przedsiębiorstwa⁵³, sądy zaś uznawały, że materiały edukacyjne udostępniane przez producenta zawierają wystarczające wyjaśnienia. Obliczenia są jednak nieprzejrzyste, a ani sprawca i jego obrońca, ani sąd nie mogą się z nimi zapoznać. Z tego względu wręcz zarzuca się władzom, że doszło do „prywatyzacji wyrokowania”⁵⁴.

6. Podsumowanie

Narzędzia oceny ryzyka powrotu do przestępstwa wykorzystujące sztuczną inteligencję przedstawiane są przede wszystkim jako obiektywne i neutralny mechanizm sprawiedliwego traktowania. Mają bowiem przyspieszyć rozpoznanie spraw i zastąpić sędziów w procesie podejmowania (uznaniowych) decyzji dzięki temu, że dostarczą na potrzeby postępowania karnego analizę uporządkowanych danych, które opierają się na szablonach podejmowania decyzji, strukturach ilościowych i metodologii nauk społecznych.

Nie ulega zatem wątpliwości, że instrumenty takie powinny spełniać wymogi charakterystyczne dla godnej zaufania sztucznej inteligencji, a także odpowiadać konstytucyjnym standardom dotyczącym prawa do sądu oraz przepisom prawa karnego materialnego. Analiza orzecznictwa sądów amerykańskich i przegląd uwag krytycznych zgłaszanych

⁵² E.A. Płocha: *O pojęciu sztucznej inteligencji...*, s. 281–283.

⁵³ W tym zakresie należałoby ustalić, czy ujawnienie tych informacji nie pozwoli na manipulowanie wynikami oceny ryzyka. Kwestionariusz zawiera bowiem także pytania dotyczące ocen.

⁵⁴ Zob. A. Nishi: *Privatizing sentencing: A delegation framework for recidivism risk assessment*. „Columbia Law Review” 2019, vol. 119, no. 6. <https://columbialawreview.org/content/privatizing-sentencing-a-delegation-framework-for-recidivism-risk-assessment/> [dostęp: 2.09.2023].

wobec aplikacji COMPAS ujawniają, że wskazane warunki nie zostały dotąd spełnione. Mimo to trzeba pamiętać, że narzędzia oceny ryzyka recydywy są nadal udoskonalane i może w przyszłości staną się godne zaufania, tj. zgodne z prawem, etyczne i solidne. Stąd też należy się zastanowić nad ich umiejscowieniem w szeroko rozumianym procesie karnym. Dotychczasowe doświadczenia organów amerykańskiego systemu wymiaru sprawiedliwości pozwalają wyodrębnić zagadnienia, które powinny zostać zweryfikowane i rozstrzygnięte przed wdrożeniem takich rozwiązań technicznych.

Po pierwsze, trzeba by ustalić, jak wyniki analizy predykcyjnej oddziałują na podejmowanie decyzji procesowych w zależności od przypisania omawianym instrumentom określonego miejsca w prawie dowodowym. Chodzi o wskazanie, czy algorytmy te mają funkcjonować jako narzędzia badawcze stosowane przez biegłych sądowych (przy opracowaniu opinii bądź ekspertyz) lub kuratorów sądowych (na potrzeby sporządzenia wywiadu środowiskowego), mogące wpływać pośrednio na rozstrzygnięcie sprawy. A może powinny stanowić odrębny (samodzielny) rodzaj dowodu, mogący bezpośrednio wpływać na wynik sprawy? Dowód taki przygotowywaliby specjalista lub biegły, którzy obsługiwaliby program i udzielali dodatkowych wyjaśnień, lub też sam organ procesowy generowałby raporty i załączał do akt sprawy. Po drugie, należałoby rozważyć, jaką pozycję przyznać ocenom ryzyka recydywy dokonywanym przez sztuczną inteligencję i w jaki sposób można je kwestionować. Uzasadnione obawy budzi bowiem przeprowadzanie przeciwdowodów, co unaoczniał m.in. przebieg postępowania w sprawie *State vs. Loomis*, w którym dopuszczono dowód z opinii biegłego podważającego możliwość posłużenia się aplikacją COMPAS, a jednocześnie, mimo podniesionych istotnych wątpliwości, odmówiono przesłuchania przedstawicieli producenta oprogramowania. Po trzecie, niezbędne byłyby zmiana przepisów prawa w zakresie obowiązkowych pouczeń i sposobu formułowania uzasadnień orzeczeń oraz edukacja uczestników postępowania karnego.

Niezależnie od zaprezentowanych tu rozważań należy wreszcie zasygnalizować, że wprowadzenie omawianych rozwiązań technicznych ma służyć eliminacji lub przynajmniej ograniczeniu błędów ludzkich. Tymczasem podstawą algorytmicznej oceny ryzyka recydywy są dane, które gromadzą ludzie w aktach spraw karnych, a raport generowany przez sztuczną inteligencję podlega swobodnej ocenie dowodów i nie wiąże organu procesowego. Stąd trudno jest uznać, że narzędzia te mogą zrealizować stawiany przed nimi cel, nawet jeżeli w literaturze przedmiotu postuluje się system mieszany. Uzasadnione obiekcje wzbudzają też wydatki na ich wdrożenie, utrzymanie, ciągłe monitorowanie i aktu-

alizowanie, które wydają się niewspółmierne do korzyści wynikających ze stosowania takich programów. Co istotne, nasuwa się obawa, że ich używanie może stworzyć barierę utrudniającą prowadzenie obrony, gdy znajdzie potrzeba zakwestionowania szacowań algorytmu – wystąpi opór przed dopuszczeniem przeciwdowodu związany m.in. z koniecznością pozyskania biegłego, czasochłonnością i kosztochłonnością tej czynności procesowej.

Bibliografia

Literatura

- Burdziak K., Rüütel E.: *Prognoza kryminologiczna. Porównanie rozwiązań estońskich i polskich*. Lublin 2022.
- Criminal law – Sentencing guidelines – Wisconsin Supreme Court requires warning before use of algorithmic risk assessments in sentencing – State v. Loomis*, 881 N.W.2d 749 (Wis. 2016). „Harvard Law Review” 2017, vol. 130, issue 5, s. 1530–1537.
- Dressel J., Farid H.: *The accuracy, fairness, and limits of predicting recidivism*, „Science Advances” 2018, vol. 4, issue 1, s. 1–5. <https://doi.org/10.1126/sciadv.aao5580>.
- Eckhouse L. et al.: *Layers of bias: A unified approach for understanding problems with risk assessment*. „Criminal Justice and Behavior” 2019, vol. 46, issue 2, s. 185–209. <https://doi.org/10.1177/0093854818811379>.
- Flores A.W., Bechtel K., Lowenkamp C.: *False positives, false negatives, and false analyses: A rejoinder to „Machine bias: There’s software used across the country to predict future criminals. And it’s biased against blacks”*. „Federal Probation” 2016, vol. 80, no. 2, s. 38–46.
- Kowalewska-Łukuć M., Burdziak K.: *The use of AI in formulating a criminological prognosis of an offender*. „Technium Social Sciences Journal” 2021, vol. 25, s. 115–125.
- Mamak K.: *Prognoza kryminologiczna*. W: Idem: *Rewolucja cyfrowa a prawo karne*. Kraków 2019.
- Nishi A.: *Privatizing sentencing: A delegation framework for recidivism risk assessment*. „Columbia Law Review” 2019, vol. 119, no. 6. <https://columbialawreview.org/content/privatizing-sentencing-a-delegation-framework-for-recidivism-risk-assessment/> [dostęp: 2.09.2023].
- Płocha E.A.: *O pojęciu sztucznej inteligencji i możliwościach jej zastosowania w postępowaniu cywilnym*, „Prawo w Działaniu” 2020, nr 44, s. 273–291.
- Schwerzmann K.: *Abolish! Against the use of risk assessment algorithms at sentencing in the US criminal justice system*. „Philosophy & Technology” 2021, no. 34. <https://doi.org/10.1007/s13347-021-00491-2>.
- Stańdo-Kawecka B.: *Korzystanie z informacji o wynikach szacowania ryzyka i potrzeb przez sąd wyrokujący*. W: Eadem: *Polityka karna i penitencjarna między punitywizmem i menedżeryzmem*. Warszawa 2020, s. 244–245.

- Stańdo-Kawecka B.: *Ruch „What Works” i „technologie szacowania ryzyka”*. W: Eadem: *Polityka karna i penitencjarna między punitywizmem i menedżeryzmem*. Warszawa 2020, s. 227–235.
- Stańdo-Kawecka B.: *Wybrane problemy profesjonalizacji organów probacyjnych i klasyfikacji sprawców oddanych pod dozór do grup ryzyka*. „Nowa Kodyfikacja Prawa Karnego” 2014, t. 33.
- Wójcik D.: *Stosowanie w postępowaniu karnym narzędzi diagnostyczno-prognostycznych służących oszacowaniu ryzyka powrotności do przestępstwa*. „Prawo w Działaniu. Sprawy karne” 2013, nr 16, s. 59–102.

Akty prawne

- Konstytucja Rzeczypospolitej Polskiej z dnia 2 kwietnia 1997 r. Dz.U. 1997, nr 78, poz. 483 ze sprost. i zm.
- Ustawa z dnia 6 czerwca 1997 r. – Kodeks karny. T.j. Dz.U. 2022, poz. 1138 ze zm.
- Ustawa z dnia 6 czerwca 1997 r. – Kodeks karny wykonawczy. T.j. Dz.U. 2023, poz. 127 ze zm.
- Ustawa z dnia 6 czerwca 1997 r. – Kodeks postępowania karnego. T.j. Dz.U. 2022, poz. 1375 ze zm.
- Ustawa z dnia 20 lutego 2015 r. o zmianie ustawy – Kodeks karny oraz niektórych innych ustaw. Dz.U. 2015, poz. 396.
- Rozporządzenie Ministra Sprawiedliwości z dnia 26 lutego 2013 r. w sprawie sposobu wykonywania obowiązków i uprawnień przez kuratorów sądowych w sprawach karnych wykonawczych. Dz.U. 2013, poz. 335.
- Rozporządzenie Ministra Sprawiedliwości z dnia 3 czerwca 2016 r. w sprawie sposobu i trybu wykonywania czynności przez kuratorów sądowych w sprawach karnych wykonawczych. Dz.U. 2016, poz. 969 ze zm.

Orzecznictwo

- Wyrok Trybunału Konstytucyjnego z dnia 16 grudnia 2014 r. U 2/14. OTK-A 2014, nr 11, poz. 119.
- Diaz v. New York State Board of Parole. Źródło: <https://law.justia.com/cases/new-york/other-courts/2013/2013-ny-slip-op-23422.html> [dostęp: 2.09.2023].
- Malenich v. State, 928 N.E.2d 564, 574 (Ind. 2010).
- Morris v. New York State Department of Corrections and Community Supervision et al. Źródło: <https://law.justia.com/cases/new-york/other-courts/2013/2013-ny-slip-op-50604-u.html> [dostęp: 2.09.2023].
- Partee v. Evans, 40 Misc.3d 896.
- Rayner v. New York State Department of Corrections and Community Supervision. Źródło: <https://law.justia.com/cases/new-york/other-courts/2023/2023-ny-slip-op-23293.html> [dostęp: 23.01.2024].
- State v. Loomis, 881 N.W.2d 749 (Wis. 2016).
- State v. Samsa, 2015 WI App 6, 359 Wis.2d580, 859 N.W.2d149.
- Williams v. New York, 49 Misc.3d 732.

Źródła internetowe

- Angwin J. et al.: *Machine bias. There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica, 23.05.2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [dostęp: 2.09.2023].
- Brennan T., Dieterich W., Mendoza Ch.: *COMPAS risk scales: Demonstrating accuracy equity and predictive parity*. Performance of the COMPAS risk scales in Broward County, 8.07.2016. Źródło: https://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf [dostęp: 25.08.2023].
- Caher J.: *Effect of risk assessment rule on parole decisions is unclear*. 30.04.2012. Źródło: <https://www.cherylkatesesq.com/news/2018/9/6/effect-of-risk-assessment-rule-on-parole-decisions-is-unclear-april-30-2012-by-john-caher> [dostęp: 30.08.2023].
- Equivant: *Practitioner's guide to COMPAS Core*. 4.04.2019. <https://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf> [dostęp: 2.09.2023].
- Equivant: *What is COMPAS?*. <https://www.equivant.com/faq/> [dostęp: 25.08.2023].
- Lansing S.: *New York State COMPAS – probation risk and need assessment study: Examining the recidivism scale's effectiveness and predictive accuracy*. Division of Criminal Justice Services, 2012. https://www.criminaljustice.ny.gov/crimnet/ojsa/opca/compas_probation_report_2012.pdf [dostęp: 2.09.2023].
- Larson J. et al.: *How we analyzed the COMPAS recidivism algorithm*. ProPublica, 23.05.2016. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> [dostęp: 2.09.2023].
- Northpointe COMPAS risk assessment. <https://assets.documentcloud.org/documents/2839240/Sample-Risk-Assessment-COMPAS-Results.pdf> [dostęp: 1.09.2023].
- Petition to the Supreme Court of the State of New York. https://law.yale.edu/sites/default/files/documents/pdf/mfia_rayner_nys.pdf [dostęp: 2.09.2023].
- Public Safety Risk Assessment Clearinghouse: History of risk assessment*. Źródło: strona WWW Bureau of Justice Assistance, U.S. Department of Justice: <https://bja.ojp.gov/program/psrac/basics/history-risk-assessment#gcov4e> [dostęp: 25.08.2023].
- Rayner M.: *Commentary: This algorithm is used to deny inmates parole. We're not allowed to know anything about it*. „Times Union” [online], 5.05.2023. <https://www.timesunion.com/opinion/article/commentary-proprietary-algorithm-deny-inmate-18001304.php> [dostęp: 2.09.2023].
- Taylor A. „Mac”: *AI prediction tools claim to alleviate an overcrowded American justice system... but should they be used?*. „Stanford Politics” [online], 13.09.2020. <https://stanfordpolitics.org/2020/09/13/ai-prediction-tools-claim-to-alleviate-an-overcrowded-american-justice-system-but-should-they-be-used/> [dostęp: 25.08.2023].

- U.S. Department of Justice. Office of the Attorney General: *The First Step Act of 2018: Risk and needs assessment system*. <https://www.ojp.gov/First-Step-Act-of-2018-Risk-and-Needs-Assessment-System> [dostęp: 2.09.2023].
- Yong E.: *A popular algorithm is no better at predicting crimes than random people*. „The Atlantic” [online], 17.01.2018. <https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/> [dostęp: 2.09.2023].