Nowoczesne Systemy Zarządzania
Zeszyt 18 (2023), nr 4 (październik-grudzień)
ISSN 1896-9380, s. 49-68
DOI: 10.37055/nsz/188842

Instytut Organizacji i Zarządzania
Wydział Bezpieczeństwa, Logistyki i Zarządzania
Wojskowa Akademia Techniczna
w Warszawie

Modern Management Systems
Volume 18 (2023), No. 4 (October-December)
ISSN 1896-9380, pp. 49-68
DOI: 10.37055/nsz/188842

Institute of Organization and Management
Faculty of Security, Logistics and Management
Military University of Technology
in Warsaw

Propozycja wykorzystania uczenia przez wzmacnianie w celu optymalizowania podejmowania decyzji w zakresie przeciwdziałania praniu pieniędzy oraz finansowania terroryzmu (część 2)

A proposal to use reinforcement learning to optimize decision-making in the field of counteracting money laundering and terrorist financing (Part 2)

Maciej Aleksander Kędzierski
Radca prawny, OIRP Warszawa, Polska
sulawezi.mk@onet.eu; ORCID: 0000-0003-3074-1355

Abstrakt. Uczenie przez wzmacnianie skupia się nie tylko na uczeniu pojedynczego agenta, lecz także zastosowanie tej metody znajduje swoje odzwierciedlenie w wieloagentowym działaniu. To kwestia istotna z punktu widzenia tego, że proces decyzyjny i zarządzanie informacją w systemie AML/CFT dla instytucji obowiązanej pozostaje coraz bardziej procesem skomplikowanym. W konsekwencji należy wprowadzić także, chcąc zastosować metodę uczenia przez wzmacnianie, wielość agentów zarówno w relacji ze środowiskiem, jak i w relacji ze sobą. Wobec tego rodzaju rozwiązań możliwe jest do zastosowania wieloagentowe uczenie się przez wzmacnianie czy koncepcja półniezależnej metody szkolenia polityk ze współdzieloną reprezentacją dla heterogenicznego, wieloagentowego uczenia się przez wzmacnianie. Ponadto mając na uwadze fakt, że proces decyzyjny AML/CFT czerpie jedynie pomocniczo rozwiązania ze sztucznej inteligencji, w tym systemie zarządzania niezbędny pozostaje także czynnik ludzki. Wobec tego rodzaju potrzeb jako wyjściowe rozwiązanie można wskazać Reinforcement Learning from Human Feedback, które zapewnia w uczeniu czynnik ludzki.
Słowa kluczowe: uczenie przez wzmacnianie, pranie pieniędzy, wieloagentowy, zbiór uczący, sprzężenie zwrotne

Abstract. Reinforcement learning focuses not only on teaching a single agent, but also the use of this method is reflected in multi-agent operation. This is an important issue from the point of view that the decision-making process and information management in the AML/CFT system for the obligated institution remains an increasingly complex process. Consequently, if we want to use the reinforcement learning

method, we must also introduce a multiplicity of agents both in relation to the environment and in relation to each other. Given this type of solutions, it is possible to use multi-agent reinforcement learning or the concept of a semi-independent policy training method with a shared representation for heterogeneous, multi-agent reinforcement learning. Bearing in mind the fact that the AML/CFT decision-making process only derives solutions from artificial intelligence, the human factor also remains essential in this management system. Given these types of needs, the initial solution can be Reinforcement Learning from Human Feedback, which ensures the human factor in learning.
Keywords: reinforcement learning, money laundering, multi-agent, training set, feedback

Wprowadzenie

W części pierwszej artykułu przedstawiono sposoby funkcjonowania, w ramach RL, pojedynczego agenta w reakcji ze środowiskiem. Nagroda więc, jako element sterowania, skupiała się wyłącznie na jednym agencie. Istnieją jednak także możliwości zadaniowania więcej niż jednego agenta na potrzeby uzyskiwania optymalizacji funkcjonowania i realizacji wyznaczonego zadania w procesie decyzyjnym AML/CFT w instytucjach obowiązanych. Celem drugiej części artykułu jest więc zaproponowanie w zakresie zarządzania informacją w obszarze AML/CFT także metody RL, ale z wykorzystaniem kilku agentów, oraz przedstawienie uczenia się przez wzmacnianie na podstawie informacji zwrotnych od ludzi (Reinforcement Learning from Human Feedback, RLHF). Druga część artykułu kończy krótkie rozważania poświęcone zastosowaniu metody uczenia przez wzmacnianie na potrzeby analityczne i decyzyjne instytucji zobligowanych w związku z potrzebą realizacji obowiązków ustawowych w zakresie przeciwdziałania praniu pieniędzy oraz finansowaniu terroryzmu. Jako metodę badawczą w części drugiej zastosowano, podobnie jak w części pierwszej, przegląd literatury oraz analizę przepisów prawnych.

1. Rozwiązania wieloagentowe (Multi-agent RL – MARL)

Wieloagentowe uczenie się ze wzmacnianiem (ang. Multi-Agent Reinforcement Learning, MARL), jako poddziedzina RL łącząca RL i MAS (ang. Multi-Agent System), dotyczy oceny, badania, w jaki sposób wielu agentów wchodzi w relację ze środowiskiem, zwłaszcza dynamicznym, oraz między sobą. Bada interakcję wielu agentów we wspólnym środowisku. MARL podejmowane jest wtedy, gdy środowisko ma charakter złożony, wieloaspektowy, wymagający współdziałania ze sobą agentów w realizacji wyznaczonego celu, a jednocześnie działań wymagających kooperacji pozytywnej. Przykładem MAS, czyli systemów wieloagentowych, jest np. wyszukiwanie informacji z sieci internetowej. Dlatego też MARL możliwe jest do zastosowania wobec zbioru klientów, a nie pojedynczego klienta lub określania trendów zwiększających ryzyko ML/FT w IO. System nagród wobec agentów powinien kształtować ich stały i zwiększany czasowo udział zwłaszcza w związku

z nieprzewidywalnością długości utrzymywania relacji klient – IO i potrzebą wykorzystania MARL w trakcie monitoringu wielości i stopnia skomplikowania transakcji danego klienta. Dlatego też pozytywne nagrody traktuje się jako zachęty dla agenta do dłuższego pozostania w grze i podejmowania takich działań, aby spróbować zdobyć wiele pozytywnych nagród (ich kumulację).

Agent jest aktywny, jeśli jego działanie nie jest spowodowane jedynie odpowiedzią na dynamizm jego środowiska i może on wykazać się celowym zachowaniem oraz przejąć inicjatywę w stosownych przypadkach. Celem interakcji jest uzupełnienie rozwiązywania problemów, które powierzono agentowi, a także pomoc innym graczom otoczenia w ich działalności (Bartuś, 2013, s. 39). Obszary środowiska, możliwości funkcjonowania agentów w IO to: znany już obszar AML/CFT, obszar *compliance* oraz obszar działań antyfraudowych. Mogą być to środowiska niezależne, a jednocześnie wymagające współdziałania aktywizujących się w nich agentów. Wydaje się, że w tym przypadku o uplasowaniu agentów w poszczególnych obszarach i ich rodzaju oraz zakresu zadań mogą decydować (pierwotnie) polityki bezpieczeństwa IO. Na potrzebę powiązania środowisk AML/CFT i fraudowego wskazuje zwłaszcza taktyka sprawców, co w przeciwdziałaniu kontrwykrywczo musi posiadać swoje umotywowanie we współdziałaniu agentów. W przypadku legalizowania dochodów niekiedy wystarczające jest jedynie wykonywanie zleceń na różnych instrumentach finansowych i usługach oferowanych przez IO. Oznacza to, że zacieranie śladów przestępczego źródła środków może odbywać się z użyciem tych samych aktywów, ale z wielością transakcji (warstwowanie). Dotyczy to także wprowadzenia pustych transakcji (bez celu finansowo-gospodarczego) lub fasadowych transakcji (zlecanych jedynie po to, aby ich ilość powodowała pozorację działań gospodarczych). Z kolei w przypadku fraudów sprawcy posługują się dodatkowo różnego rodzaju fałszywymi, podrobionymi lub przerobionymi „dokumentami" stanowiącymi także pretekst do zleceń transakcyjnych. Wykrywanie oszustw, opierając się na obrazach, polega głównie na analizie obrazów związanych z transakcjami finansowymi, takimi jak banknoty, czeki, faktury, paragony itp., co w konsekwencji powoduje, iż próbuje się wykryć i ustalić, czy są one fałszywe lub naruszone. Wykrywanie oszustw tekstowych zajmuje się głównie analizą informacji tekstowych związanych z transakcjami finansowymi, takimi jak zapisy transakcji, profile klientów, informacje zwrotne, recenzje itp., i przeciwdziałanie stara się wykryć, czy są one niespójne lub nietypowe (Tong, Shen, 2023). Dlatego też na potrzeby AML/CFT działania przeciwfraudowe znacznie „ubogacają" zbiory uczące dla agenta działającego w środowisku AML/CFT. O ile efekt dla CFT jest już nagradzany, gdy agent wykryje tożsamość klienta z osobą wyznaczoną na liście sankcyjnej, to ze środowiska fraudów możliwa jest także konfrontacja nie tylko osoby, lecz także jej uplasowania w dokumentach, posiadanych banknotach, fakturach czy czekach, które posłużyły do wyłudzenia środków (wykrycie obrazowe i tekstowe). Stąd kooperacja agentów w tych środowiskach pozostaje niezbędna dla ujawnienia i utrwalenia całości przestępczego procederu.

Podejście takie można uznać za metodę określania procederu prania pieniędzy czy finansowania terroryzmu przez badanie kreowania przestępstw źródłowych jako aktywności pierwotnych wobec wtórnych procederów ML/FT.

Możliwe jest więc zaistnienie takich sytuacji, w jakich agenci będą występowali przy takich defektach (czasami uzasadnionych) zarządzania bezpieczeństwem, w których będziemy mieli do czynienia z fragmentarycznością informacji lub brakiem umiejętności agenta do samodzielnej realizacji zadania, co oznaczałoby, że każdy z agentów będzie miał ograniczony horyzont postrzegania zadania. Ale także w przypadku gdy występuje wąska lub szeroka autonomia poszczególnych agentów – w zależności od obranej architektury systemu agentowego, rozproszenia danych (ich decentralizacja) oraz funkcji agentów – czy asynchroniczność działania poszczególnych agentów (Bartuś, 2013, s. 41; Weyns, 2010). Należy jednak zaznaczyć, że działania agentów będą nakierowane nie na usprawnienie funkcjonowania IO, lecz na wykrywanie zagrożeń/niebezpieczeństw związanych z naruszaniem prawa zwłaszcza przez klientów IO. Ponadto należy mieć na uwadze również specyfikę uzyskiwania danych zasilających zbiory uczące dla multiagentów. W konsekwencji dane te są generowane od podmiotów świadomie niejednokrotnie posługujących się danymi błędnymi lub zmanipulowanymi, aby wywołać w IO fałszywe pozytywne przeświadczenie o ich prawdziwości. Wobec tych zagrożeń stosuje się metody – przeciwstawnego uczenia maszynowego (ang. Adversarial Machine Learning, AML). Konsekwencją jest to, aby w danych treningowych umieszczać przykłady kontradyktoryjne, które zostały celowo zmanipulowane w celu oszukania modelu sztucznej inteligencji. Podczas szkolenia, wystawiając ów model, trzeba zwrócić uwagę agentowi na te przykłady, ponieważ pomoże to jemu nauczyć się rozpoznawać i bronić się przed podobnymi atakami w przyszłości (Frąckiewicz, 2023).

Wieloagentowe uczenie się przez wzmacnianie (MARL) można uznać za rozszerzenie RL, które uwzględnia interakcje między wieloma agentami w zmieniającym się środowisku. Nagroda skalarna (ang. *scalar reward*), tzn. wielkość opisywana skalarem w sensie matematycznym sygnału, oceniana jest jako jakość każdego przejścia, a agent musi zmaksymalizować skumulowaną nagrodę w trakcie interakcji. Tym samym agenci muszą nauczyć się dostosowywać swoje działania nie tylko w odniesieniu do zmian zachodzących w środowisku, lecz także do zmian w zachowaniu innych agentów. Obszar funkcjonowania w takim przypadku jest znaczne szerszy i może obejmować współpracę, konkurencję czy kooperację pomiędzy poszczególnymi agentami. Konkurencyjne środowiska obejmują między innymi nagrody o sumie zerowej, polegające na podwyższeniu nagrody jednego agenta, co bezpośrednio zmniejsza nagrodę innego agenta. Środowiska kooperacyjne obejmują nagrody, które są w całości wspólne wszystkim agentom. Najwcześniej wyróżnia się działania oparte na wartościach i na polityce.

Ponadto należałoby zaznaczyć, że działanie multiagentowe może być realizowane zarówno jako sposób na prowadzenie procederu prania pieniędzy czy finansowania terroryzmu (ML/FT), jak i metoda na przeciwstawienie się takiej aktywności sprawców w ramach działań AML/CFT podejmowanych przez IO. Tym samym poszczególni agenci będą mogli być zadaniowani jednorodzajowo lub każdy z nich pod innym kątem na podstawie zasad, ustalonego modelu czy wartości. Tym samym IO powinna pre-działaniowo ocenić obszary funkcjonowania agenta, tak aby jego rolę w całości procesu AML/CFT właściwie zidentyfikować i nazwać z zachowaniem kooperacji pomiędzy nimi. Oznacza to, że inaczej można definiować akcje agenta w zakresie powtarzalności procesów i badania odchyleń od ustalonego wzorca (w kierunku odtworzeniowym), a inaczej zadaniować innego agenta w kierunku poszukiwania nowych procederów ML/FT reprezentowanych przez sprawców (w kierunku predykcyjnym). Wspólną podstawą do tych działań (część wspólna danych uczących) będzie efekt stosowania środków bezpieczeństwa finansowego w postaci identyfikacji i weryfikacji oraz dotychczasowych „doświadczeń" agentów z akcji prowadzonych w środowisku IO. Tym samym będzie można wykreować „efekt przeciwdziałania" i „efekt postępowania" podmiotów identyfikowanych z ML/FT. Będą to więc stany pożądane wobec oceny kontrreakcji i jej odpowiedniego zastosowania w ramach prowadzonej oceny ryzyka. Zadaniowanie agentów będzie także związane z tym, czy IO będzie chciała dokonać weryfikacji modelu postępowania z klientem, czy działania będą epizodyczne, długofalowe oraz czy planuje się powtarzalność tych samych działań, czy poszerzanie o nowe działania. Ponadto należy zwrócić uwagę na to, że ML nie musi w IO dotyczyć wyłącznie procedur AML/CFT. Aktywność agenta może także odnosić się do działań o charakterze komercyjnym (marketingowym), co może być wykorzystane na potrzeby zasilania zbiorów uczących dla agentów funkcjonujących na odcinku AML/CFT, a także w kwestiach, takich jak oszustwa, wyłudzanie kredytów czy działanie na szkodę innego klienta. W celu budowania baz danych IO możliwe jest wykorzystanie informacji pochodzących z „zelektronizowanych" relacji z klientem, np. gdy ich źródłem są formularze wypełniane elektronicznie, chatboty, wirtualni agenci, procedury zatwierdzeniowe usług/produktów i ich zmian, wewnętrzna poczta elektroniczna, korespondencja e-mail, zlecenia przelewów, opłaty rachunków i kontrola wydatków czy zmiana danych kontaktowych dokonywana przez klienta za pomocą bankowości elektronicznej.

W zakresie AML/CFT metoda MARL pozostaje o tyle istotna, że w przypadku więcej niż jednego agenta poszczególni agenci mogą być przypisani do różnych środowisk lub do różnych zadań w tym samym środowisku, a jednocześnie mogą być związani potrzebą kooperacji pozytywnej wobec siebie. Takie podejście znacznie wzmacnia podbudowę czasową i merytoryczną na potrzeby budowania całościowego obrazu podejrzalności wobec zachowań klienta/transakcji w danym środowisku. Mimo że występują trzy odmiany MARL, to w omawianym zakresie najbardziej istotna jest ta, która dotyczy współpracy agentów wobec realizacji celu.

Niemniej jednak i w tym zakresie wyodrębnia się określone trudności. Zalicza się do nich to, że przestrzeń działania zmienia się wykładniczo wraz z liczbą agentów, co może powodować problemy ze skalowalnością, zjawisko znane jest jako kombinatoryczna natura MARL. W związku z tym, że każdy agent ma ograniczony dostęp do obserwacji innych, prowadzić to może prawdopodobnie lokalnie do nieoptymalnych reguł podejmowania decyzji. W środowisku w pełni współpracującym wszyscy agenci zwykle dzielą wspólną funkcję nagradzania $R^1 = R^2 = \cdots = R^N = R$. Mając na uwadze ten model, funkcja wartości i funkcja Q są identyczne dla wszystkich agentów, co umożliwia w ten sposób jednoagentowe algorytmy RL, np. umożliwia zastosowanie aktualizacji Q-learning, jeżeli wszyscy agenci są skoordynowani jako jeden decydent. Oprócz modelu wspólnej nagrody jest inny model – spółdzielnia MARL, która bierze pod uwagę średnią nagrodę zespołową. Dla każdego konkretnego agenta mogą występować różne funkcje nagród, które mogą być prywatne (specjalne) dla każdego z nich, podczas gdy celem współpracy jest optymalizacja długoterminowej nagrody odpowiadającej średniej nagrodzie:

$$\bar{R}(s,a,s') := N^{-1} \cdot \sum_{i \in N} R^i(s,a,s') \text{ dla każdego } (s,a,s') \in \mathcal{S} \times \mathcal{A} \times \mathcal{S}$$

Model średniej nagrody, który pozwala na większą heterogeniczność wśród agentów, wymaga również włączania protokołów komunikacyjnych do MARL oraz analizy wydajności algorytmów MARL (Zhang, Yang, Bas, 2021, s. 9). Działaniem MARL powinno być maksymalne kumulowanie nagród, nawet kosztem nagród pośrednich. Te zaś mogą być stanem oceny dla agenta, że postępuje we właściwym kierunku, ale nie powinien z niego rezygnować i rozumieć także „dowartościowanie", korzystając z części wspólnej dostępnej dla pozostałych agentów.

Dla przykładu, synchroniczna architektura rozproszona opiera się na architekturze jednowątkowej i wykorzystuje wiele instancji środowiska działających w procesach równoległych do jednoczesnego zbierania partii doświadczeń. Każde środowisko synchronicznie współdziała ze wspólną polityką, generując sekwencje stanów, akcji, nagród i kolejnych obserwacji. Ta partia sekwencji służy do aktualizowania wag funkcji zasad i wartości przez opadanie gradientu. Wykorzystując równoległość, synchroniczna architektura rozproszona umożliwia wydajniejsze gromadzenie danych i szybsze aktualizacje, co prowadzi do przyspieszonego szkolenia i lepszej konwergencji w uczeniu się przez wzmacnianie (Mehta, Mahajan, Kumar, 2023, s. 3-4). W przypadku natomiast działania w asynchronicznej, rozproszonej architekturze szkoleniowej w stylu IMPALA działanie wielu aktorów (agentów) następuje równolegle i asynchronicznie, wchodząc w interakcje z odpowiednimi środowiskami (Mehta, Mahajan, Kumar, 2023, s. 3-4). W tym zakresie działa także kilka komponentów:

- **pętla środowiskowa** – proces pętli środowiskowej oddziałuje ze środowiskiem, korzystając z dostępnej polityki, i dodaje zebrane doświadczenie do

bufora powtórek. Uruchamianych jest wiele równoległych procesów pętli środowiska, każdy z własną kopią pliku parametru środowiska i polityki. Aby zachować synchronizację parametrów polityki z procesem uczenia, parametry są okresowo pobierane z procesu uczącego. Wybór kroku akcji jest zoptymalizowany z użyciem automatycznej wektoryzacji, aby wybrać akcję dla wszystkich agentów w środowisku;

- **uczenie** – w tym procesie odbywa się rzeczywiste uczenie się zasad. Agent-uczeń pobiera doświadczenie z bufora powtórek i wykonuje krok optymalizacji zasad i parametry funkcji wartości. Używa się tu *pmap* (polecenie *pmap* dołącza mapę pamięci procesu lub programu. Możliwe jest wykorzystanie tej funkcji do monitorowania ruchu pamięci w określonym przedziale czasu) do automatycznego skalowania kroku optymalizacji do wielu procesorów graficznych i automatycznej wektoryzacji opartej na *vmap* (numeryczne opracowanie kartograficzne złożone z obiektów typu: punkt, linia, obszar i ich odmian, dla których współrzędne zostały zapisane w bazie danych, natomiast obraz mapy jest generowany w zależności od ustawionej skali, tak aby nie występowało zjawisko pikselizacji) (Wikipedia, 2023), co umożliwi wykonanie kroku optymalizacji dla wszystkich agentów równolegle;
- **bufor powtórek** – wszyscy agenci dodają doświadczenie do tego serwera, a osoba ucząca się pobiera próbki doświadczenia z serwera w celu optymalizacji pod kątem zasad i funkcji wartości parametru.

Ponadto zastosowanie wspólnego serwera wnioskowania, który jest używany w asynchronicznej rozproszonej architekturze (poszczególne węzły/urządzenia obliczeniowe komunikują się i synchronizują za pośrednictwem wspólnej sieci) do szkolenia RL, powoduje, iż architektura jest dalej ulepszana w celu scentralizowania procesu wnioskowania. W tej konfiguracji wielu agentów asynchronicznie wchodzi w interakcje z odpowiednimi środowiskami, zbierając trajektorie doświadczeń, tak jak poprzednio. Jednak zamiast tego, iż każdy agent przeprowadza własne wnioskowanie, agenci wysyłają zebrane doświadczenia do wspólnego serwera wnioskowania. Wykorzystując wspólny serwer wnioskowania, można osiągnąć kilka korzyści. Po pierwsze, zmniejsza to obciążenie obliczeniowe poszczególnych agentów, ponieważ nie muszą już wykonywać zadań własnego wnioskowania. Dzięki temu agenci mogą skupić się na gromadzeniu danych, co daje zwiększoną sprawność i szybszą interakcję z otoczeniem. Po drugie, sieć zapewnia korzystanie ze wspólnej polityki, która powoduje, że wszyscy agenci podejmują decyzje na podstawie tego samego zestawu parametrów (Mehta, Mahajan, Kumar, 2023, s. 7). Wskazane możliwości pozwalają na koordynowanie zadaniowania agentów wobec uzyskanej uprzednio wiedzy n-agentów działających jako agenci monitoringu wobec n-klientów oraz samouczenie i wspólne ubogacanie podstaw do wzmocnienia przed kolejnym krokiem akcji. Przedstawione założenia mogą stanowić podstawę do kompleksowego i rozwojowego zarządzania negatywną informacją w IO.

W takim przypadku kreowana wewnętrzna polityka bezpieczeństwa IO mogłaby zapewnić dysponowanie danymi dwóch kategorii: stałych zmiennych identyfikacyjnych i weryfikacyjnych oraz zmiennych danych kreowanych na bazie stosunków gospodarczych z klientem i zmiennych czynników wpływających na ryzyko indywidualne klienta. O ile taki model jest kwantyfikowany dla bezpieczeństwa AML/CFT i wobec stosowania środków bezpieczeństwa finansowego, to nie jest on obligatoryjny dla innych kategorii środowiska, tj. compliance i fraudowego. Stąd też, gdy IO wybierze ten model MARL, będzie musiał uznać dane stałe przy modyfikowaniu danych zmiennych sfokusowanych na nieprawidłowościach związanych z zapewnieniem zgodności działalności IO z przepisami o przeciwdziałaniu praniu pieniędzy/finansowaniu terroryzmu oraz oszustw.

Współdziałanie więcej niż jednego agenta jedynie w środowisku AML/CFT możliwe byłoby do wykorzystania w przypadku, gdy zależy nam na uzyskaniu wyniku określonego w art. 86 ustawy o p.p.p.f.t. (umownie: stan 86), a nie tylko wskazanego w art. 74 tej ustawy (umownie: stan 74). Oznacza to, że każdy z agentów może doprowadzić do wygenerowania SAR, ale jedynie współdziałanie, niekiedy kosztem rezygnacji z wygenerowania SAR, doprowadzić może do większej maksymalizacji wyniku w postaci skierowania do jednostki analityki finansowej informacji o uzasadnionym podejrzeniu, jakiej wymaga spełnienie przesłanek do wstrzymania transakcji lub blokowania rachunku (ten sam efekt możliwy jest do uwzględnienia w sytuacji art. 41 ust. 2 ustawy o p.p.p.f.t.) (stan 86). Jest to więc sytuacja, gdy stan 86 nie jest możliwy do wygenerowania przez pojedynczego agenta. Jest on do osiągnięcia jedynie w wyniku wzajemnej współpracy, także wówczas, gdy jeden z agentów jest już zaspokojony uzyskaniem stanu 74. W konsekwencji bez uogólnień agenci przeszkoleni nie mogą osiągnąć optymalnej polityki w nowym scenariuszu oceny. Stratą będzie to, że zaspokojony agent uzna osiągnięcie stanu 74 i odebranie nagrody jako kres swojej aktywności, a agent, który tego stanu nie osiągnął, będzie dążył do jego uzyskania. W konsekwencji żadnemu z agentów nie będzie zależało na uzyskaniu stanu 86 (w pojedynkę), który będzie stanem bardziej wynagradzanym. Do rozważenia możliwe byłoby łączenie wskazanych stanów jako „następczych" wobec stanów „pierwotnych", np. wynikających z fraudów (stan 286, w art. 286 kodeksu karnego penalizowane jest oszustwo), identyfikujących czyny zabronione, jako źródło aktywów podlegających „wypraniu" lub przekazaniu aktywów dla środowisk terrorystycznych, zwłaszcza gdy agent działający w środowisku fraudów kooperuje z agentem w środowisku AML/CFT wobec identyfikacji powstania „przestępstwa źródłowego" w tej samej instytucji IO.

Należy wskazać, że metody wieloagentowego uczenia się przez wzmacnianie (MARL) można podzielić na dwie kategorie w zależności od poziomu centralizacji w podejmowaniu decyzji i uczeniu się (scentralizowane lub zdecentralizowane). Scentralizowane podejście do uczenia się zakłada wspólny model dla działania i obserwacji wszystkich agentów. Scentralizowana polityka odwzorowuje wspólną

obserwację wszystkich agentów na wspólne działanie oraz jest odpowiednikiem zasady MPOMDP (ang. *multi-agent partially observable Markov decision proces* – wieloagentowy częściowo obserwowalny proces decyzyjny Markowa). Główna wada tego podejścia polega na tym, że jest ono scentralizowane zarówno pod względem szkolenia, jak i wykonania oraz prowadzi do wykładniczego wzrostu przestrzeni obserwacji i działań z liczbą agentów (Gupta, Egorov, Kochenderfel, 2017). W systemach zdecentralizowanych każdy agent podejmuje decyzje i uczy się samodzielnie, bez dostępu do obserwacji, działań lub polityki innych agentów. Jednak zdecentralizowanemu uczeniu się brakuje gwarancji konwergencji z powodu niestacjonarności spowodowanej przez inne czynniki. Dlatego większość współczesnych badań MARL jest zgodna z paradygmatem scentralizowanego szkolenia i zdecentralizowanego wykonywania (ang. *centralised training and decentralised execution*, CTDE) W paradygmacie tym to agenci mają dostęp do obserwacji innych agentów podczas szkolenia, ale oddzielnie wykonują własne zasady (Zhao, Jin, Chen, Guo 2023). Metoda zmierza więc w kierunku przyjęcia uczenia się z wieloma wzmocnieniami. Centralnie wykonywany algorytm MARL działa podobnie do algorytmu pojedynczego agenta. Jednakże struktura scentralizowanego systemu wieloagentowego może mieć większą skalowalność, ale gorszą koordynację niż pojedynczy agent. Ten kompromis występuje dlatego, że scentralizowany system wieloagentowy zmniejsza sprzężenie pomiędzy jego elementami (Standen, Kim, Szabo, 2023).

Jedną z propozycji rozwiązywania problemów uczenia się wieloagentowego jest koncepcja półniezależnej metody szkolenia polityk ze współdzieloną reprezentacją dla heterogenicznego, wieloagentowego uczenia się przez wzmacnianie (ang. *a semi--independent policies training method with shared representation for heterogeneous multi-agents reinforcement learning*) (Zhao, Jin, Chen, Guo 2023). Przyjmuje się schemat współdzielenia twardych parametrów do MARL w celu zrównoważenia sprzecznych wymagań specjalizacji agentów i szybkiej konwergencji sieci. Metoda ta pozwala wygenerować ogólną reprezentację danych wejściowych i wyjściowych współdzieloną ze wszystkimi agentami. Wspólna reprezentacja ułatwia sformalizowanie danych wejściowych i wyjściowych, rozwiązując w ten sposób różnorodność problemów wejściowych i wyjściowych agentów heterogenicznych oraz ułatwia włączenie schematu współdzielenia twardych parametrów. Dzięki tej wspólnej, współdzielonej reprezentacji wejścia/wyjścia wszyscy agenci będą jednakowo traktowani po przetworzeniu wejścia/wyjścia, niezależnie od rodzaju agentów.

Jednocześnie w tej metodzie zostaje wprowadzona dodatkowa wewnętrzna nagroda dla agentów, aby zachęcić do dalszej eksploracji środowiska już na początku. W przeciwieństwie do tradycyjnych nagród wewnętrznych, które opierają się na porównaniu trajektorii, proponowana nagroda wewnętrzna opiera się na przewidywaniu nadzorowanego uczenia się i jego reprezentacji wejścia/wyjścia. Wydaje się, że w przypadku AML/CFT takie podejście pozwala na podjęcie akcji przez wielu

agentów, z zachowaniem „twardych danych podstawowych" dla każdego z nich, którymi dysponuje IO. Jednakże każdy z tych agentów pozostaje różnorodny wobec siebie (heterogeniczny), ale pozostaje zachowany wspólny cel współdziałania. Cechą pozytywną tego podejścia jest możliwość odmiennego ustanawiania agentów (niejednorodnie) w tym samym środowisku przez powielanie wspólnych reprezentacji i współdzielenie parametrów między agentami tego samego typu. Taka szeroka różnorodność zadaniowania agentów i sterowania nimi za pomocą nagród może umożliwić IO różnorodne ich zadaniowanie w zależności od zmienności środowiska, które będzie na bieżąco monitorowane. Wykonawstwo wyszukiwania zagrożeń będzie można przekazać na rzecz agentów, zwłaszcza gdy powstają podejrzenia co do tego, że przestępczy proceder może dotyczyć wielu produktów (także w chmurze) lub rozproszonych miejscowo komórek organizacyjnych IO.

Odpowiedzią na wyzwanie budowania ogólnych polityk jest także np. metoda zwana rankingową pamięcią zasad (ang. *ranked policy memory*, RPM). Według autorów ideą RPM jest utrzymanie pamięci wyszukiwania polityk podczas szkolenia dla agentów. W szczególności po każdej aktualizacji szkolenia najpierw oceniane są zasady wyszkolonych agentów. Następnie podejmowana jest klasyfikacja *polis* wyszkolonych agentów przez powrót epizodu szkoleniowego i zapisanie ich w pamięci. W ten sposób uzyskuje się różne poziomy, czyli wykonanie *polis* (polityk). Rozpoczynając odcinek, agent może uzyskać dostęp do pamięci i załadować losowo próbkowaną politykę w celu zastąpienia bieżącej polityki zachowania. Nowy zestaw zasad umożliwia agentom realizację gry samodzielnej i zbieranie zróżnicowanych doświadczeń w grze o środowisko szkoleniowe. Te zróżnicowane doświadczenia zawierają wiele nowatorskich interakcji między agentami, które mogą zwiększyć zdolność ekstrapolacji MARL, zwiększając w ten sposób wydajność uogólniania (Qiu, Ma, An et al., 2023). Należy zaznaczyć, że jeśli model dokładnie prognozuje na nowych danych, stanowi to, że może uogólniać od zestawu uczącego do zestawu testowego na potrzeby zbudowania modelu, który będzie uogólniać jak najdokładniej. Zwykle buduje się model w taki sposób, aby mógł trafnie prognozować na zestawie uczącym. Jeśli zestawy uczące i testowe mają wystarczająco dużo cech wspólnych, oczekuje się (zakłada), że model będzie również dokładnie uogólniał dane z zestawu testowego (Muller, Guido, 2023).

2. Uczenie się przez wzmacnianie na podstawie informacji zwrotnych od ludzi – Reinforcement Learning from Human Feedback (RLHF)

RLHF to stosunkowo nowe rozwiązanie, które łączy techniki uczenia się przez wzmacnianie, takie jak nagrody i porównania z ludzkimi wskazówkami, w celu wyszkolenia agenta sztucznej inteligencji (ang. Artificial Intelligence, AI). W RLHF testerzy i użytkownicy dostarczają bezpośrednich informacji zwrotnych, aby zoptymalizować

model językowy dokładniej niż samouczenie się. RLHF jest używany głównie w przetwarzaniu języka naturalnego (ang. Natural Language Processing, NLP) do rozumienia agentów AI w aplikacjach, takich jak chatboty i agenci konwersacyjni, zamiana tekstu na mowę i streszczanie (Patrizo, 2023). RLHF jest procesem iteracyjnym, ponieważ zbieranie informacji zwrotnych od ludzi i udoskonalanie modelu za pomocą uczenia się przez wzmacnianie jest powtarzane w celu ciągłego doskonalenia. W przeciwieństwie do tradycyjnego uczenia się przez wzmacnianie (RL), w którym agent uczy się metodą „prób i błędów", RLHF umożliwia szybsze i bardziej ukierunkowane uczenie się dzięki wykorzystaniu ludzkiej wiedzy. W tej konfiguracji rozwiązań metoda RLHF możliwa jest do zastosowania na poziomie AML/CFT.

Nie wszystkie informacje będące elementem wprowadzenia do zbioru uczącego możliwe są do wygenerowania wyłącznie ze zbioru danych uzyskiwanych w systemach technicznych IO. Należy zauważyć, że czynnik ludzki to wykorzystanie wiedzy eksperckiej do zarządzania informacją w IO opartą na interpretacji informacji za pośrednictwem zdolności myślenia ludzkiego. W przypadku AML/CFT może ona dotyczyć różnych zdarzeń i potrzeby ich włączenia w proces AI. Dla przykładu można wskazać potrzebę zachowania kompatybilności z operacją specjalną jednostki współpracującej, poleceniami służbowymi członka zarządu w ramach wykonywanego nadzoru nad systemami bezpieczeństwa itp. Dotyczy to więc tych zdarzeń, w których nie można nauczyć agenta modelu na podstawie danych (zwłaszcza technicznych). Rolą czynnika ludzkiego jest wykreować oraz włączyć do procesu analitycznego agenta takich cech, które nie są możliwe do wykreowania ich przez uczenie maszynowe. Ostatecznie cechy te wprowadza się do zbioru uczącego agenta. Działanie podejmuje się w ten sposób, aby wspomóc algorytm uczenia maszynowego. Można więc zakodować wcześniejszą wiedzę o naturze zadania w cechach. Dodane cechy nie zmuszą algorytmu uczenia maszynowego do korzystania z niej, zapewniają jednak „pogłębienie wiedzy" dla agenta o tę wartość dodaną, której on sam nie mógłby wykreować (Muller, Guido, 2023, s.208). Ważne jest także to, aby cecha wprowadzanego elementu nie pozostawała poza zakresem wartości cech umieszczonych w zestawie uczącym, ponieważ w razie popełnienia błędu niewykonanie tego „wartościowania" będzie niewidoczne dla agenta. Tradycyjne algorytmy uczenia maszynowego są stosunkowo słabe w przypadku złożonych relacji nieliniowych i danych wielowymiarowych. Trudno też uchwycić wielopoziomową strukturę i złożone powiązania transakcji finansowych. Dostosowanie parametrów modelu podczas procesu uczenia wymaga interwencji człowieka, co utrudnia adaptacyjną aktualizację modelu, a jednocześnie skutkuje niemożnością poradzenia sobie z nowymi oszukańczymi zachowaniami (Tong, Shen, 2023). Uwagi te jednak dotyczą środowiska fraudowego, choć można je także przyjąć odpowiednio do środowiska AML/CFT.

Czynnik ludzki dla oceny ryzyka indywidualnego, analityki komórek AML/CFT czy procedury KYC nie powinien być wyeliminowany, a wręcz musi pozostawać jako końcowy efekt procesu decyzyjnego wygenerowującego informację z IO do JAF.

W przypadku RLHF czynnik ludzki, a zwłaszcza jako element dostarczania wiedzy dla agenta pracującego w określonym środowisku, może być pomocny na bieżąco z zastosowaniem wzmocnionego (właśnie wiedzą tego czynnika) uczenia maszynowego w pośrednich etapach decyzyjnych. Różnica między RL a RLHF polega na źródle sprzężenia zwrotnego. RL opiera się na autonomicznej eksploracji, podczas gdy RLHF integruje wskazówki człowieka w celu przyspieszenia uczenia się. RLHF wykorzystuje informacje zwrotne od ludzi, aby szybciej trenować modele uczenia się przez wzmacnianie. Tym samym można oszczędzić czas, wykorzystując opinie ludzi do kierowania nauką, zamiast polegać na błędnych lub ograniczonych celach. Może to także naprawić wady i zwiększyć możliwości wyboru modelu. Ludzki mechanizm nagrody pozwala modelowi lepiej zrozumieć swoje środowisko i szybciej osiągnąć konwergencję. Tym samym czynnik ludzki pomaga, okazjonalnie dostarczając dodatkowy sygnał nagrody/kary, albo dostarcza dane potrzebne do wytrenowania modelu nagrody. Należy także zauważyć, że wskazówki, jakie może kierować czynnik ludzki na rzecz uczenia agenta, będą uzyskiwane z różnych obszarów, ale wspólnym ich mianownikiem będzie postrzeganie cech charakterystycznych dla stanów związanych z różnymi fazami prania pieniędzy czy finansowania działalności terrorystycznej. Czyli mimo ich początkowego rozproszenia będą one kanalizowane w ocenę zachowań potencjalnych sprawców jako kreatorów łańcucha przestępczego. Ponadto istotne jest, że agent otrzyma z relacji człowiek – człowiek wskazówki, których nie otrzymałby, gdyby przetwarzałby dane jedynie w środowisku „technicznym" lub „półtechnicznym" AML/CFT.

Innego rodzaju wskazówki mogą pochodzić z relacji człowiek – produkt/usługa (IO) czy z obszarów równoległych innych IO, wykonujących podobne usługi, w których zauważono „zachowania podejrzane", a których nie doświadczyła ta IO', w środowisku której działa agent. Połączenie takich wskazówek z odpowiednią gradacją nagrody dla agenta daje możliwości szybkiego rozpoznania w środowisku IO takich zachowań podejrzanych, które dotychczas były znane tylko innej IO' (np. naznaczając je jako bardziej wynagradzane niż inne „wykrycia" stanów przez agenta, stworzenie nowej funkcji nagrody). Wskazane podejścia mogą zwiększyć potencjał agenta wtedy, gdy IO podejrzewa zmowę w celu dokonania przestępstwa i potrzebę jej obserwowania, jak wytypowane osoby zmowy zachowują się wobec produktów/usług świadczonych przez IO (Abramson, Ahuja, Carnevale, Georgiev, 2022).

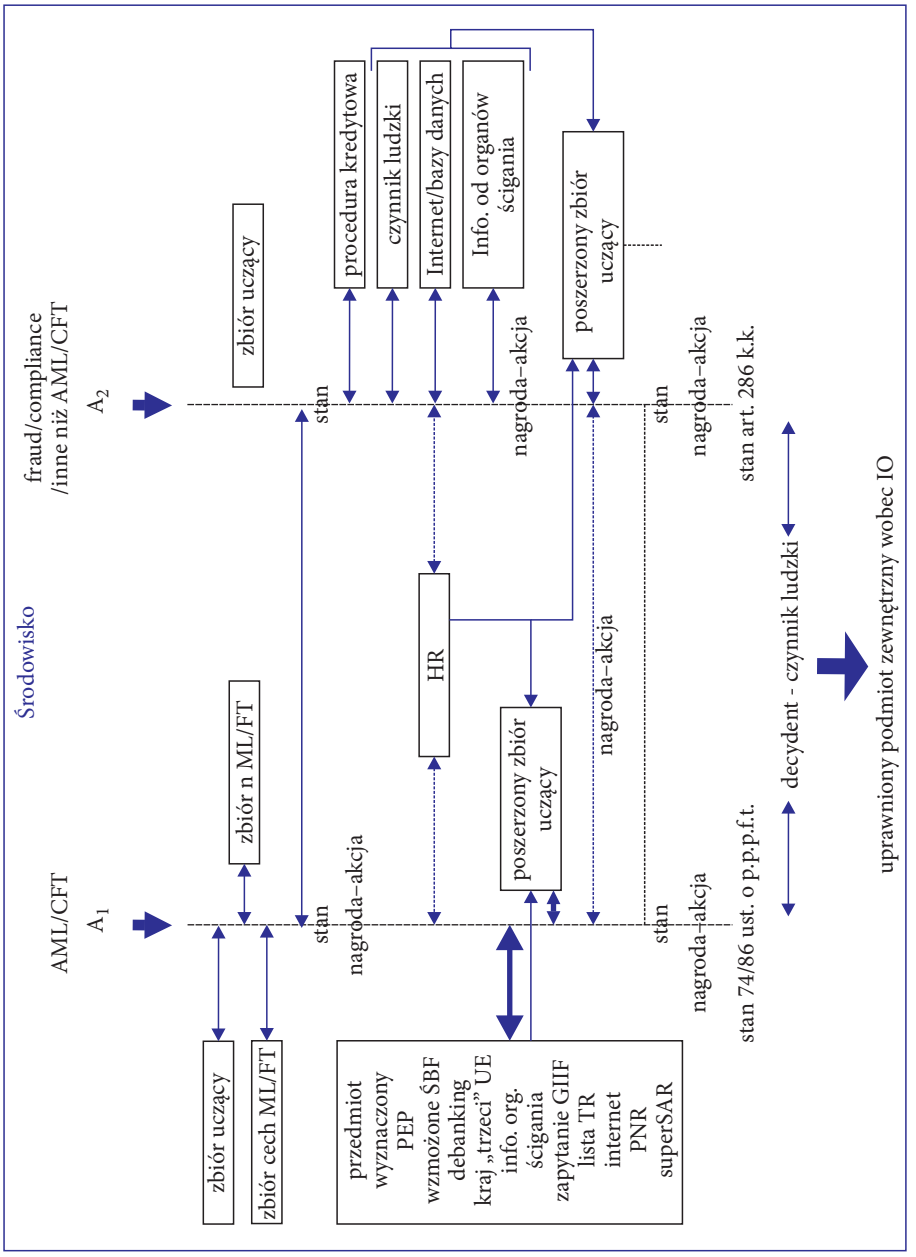
W RLHF istotne pozostaje także wskazanie rodzajów informacji zwrotnych uzyskiwanych od czynnika ludzkiego. Do nich można zaliczyć między innymi: informację o względnej jakości swoich działań (wskazująca na preferowany wybór rodzaju działania dla agenta spośród alternatywnych możliwości), modyfikację sygnału nagrody w trakcie procesu agent – środowisko, otrzymanie przez agenta informacji zwrotnej o popełnionym błędzie i wymaganej korekcie. Ponadto w całości procesu ważnym elementem jest także sposób włączenia informacji od czynnika ludzkiego na rzecz agenta (Dhaduk, 2023). W tym przypadku wymienia się techniki, takie jak uczenie

się przez naśladownictwo: agent uczy się na podstawie demonstracji prowadzonych przez eksperta-człowieka, czy odwrotne uczenie się ze wzmacnianiem (IRL): agent wnioskuje w sprawie podstawowej funkcji nagrody na podstawie demonstracji na ludziach. IRL może zapewnić wgląd w jego preferencje dotyczące ryzyka, czas trwania i inne czynniki informujące o optymalnych strategiach alokacji aktywów dostosowanych do ich preferencji, które można zastosować do alokacji aktywów, aktywne uczenie się na podstawie demonstracji (ALfD): agent wchodzi w interakcję z człowiekiem, aby aktywnie zażądać demonstracji w niektórych sytuacjach wymagających wskazówek, uczenie się przez wzmacnianie przez człowieka w pętli (HITL RL): agent wchodzi w interakcję z człowiekiem w czasie rzeczywistym, otrzymując informacje zwrotne i poprawki podczas uczenia się (Dhaduk, 2023). Istotne jest jednak to, aby nie „przegrzać" czynników uczenia się pochodzących od człowieka, zwłaszcza gdy agent sam musi dojść w „środowisku technicznego uczenia się" do wniosków, które nie zostały wykryte przez czynnik ludzki, lub gdy człowiek chce pozostawić decyzyjność agentowi, tak aby wytworzyć efekt przynależny „myśleniu agenta", a nie jako wynik „podpowiedzi" człowieka.

Przedstawione możliwości współdziałania czynnika ludzkiego i sztucznej inteligencji w obszarze AML/CFT wymuszają na pracownikach komórek IO odpowiedzialnych za przeciwdziałanie praniu pieniędzy oraz finansowaniu terroryzmu, a także za przeciwdziałanie innego rodzaju naruszeniom (np. za walkę z oszustwami na szkodę IO lub innego klienta) nauczenia się monitorowania pracy agentów oraz współdziałania z nimi, zwłaszcza w zakresie umiejętności zadaniowania, weryfikacji otrzymanych rezultatów oraz oceny wyników. Sztuczna inteligencja już odnotowuje znaczące sukcesy w wykrywaniu złożonych oszustw związanych z praniem pieniędzy. Ogromna skala współczesnych oszustw finansowych wymaga możliwości zastosowania sztucznej inteligencji i uczenia maszynowego (ML), aby zapewnić w czasie rzeczywistym najwyższą elastyczność, której brakuje tradycyjnemu wykrywaniu oszustw opartemu na regułach. Uczenie maszynowe jest zdolne do „ciągłego" uczenia się, gdy zestawy danych są wprowadzane do algorytmu ML. Wobec ogromnej liczby popełnianych oszustw te zestawy danych szybko się zmieniają, a algorytm ML reaguje i optymalizuje działania związane z wykrywaniem oszustw (Eastnets, open-source AI, 2023). Zadaniowanie agenta wymaga od pracowników komórki AML/CFT IO z nim współpracujących przyjęcia takiego przekazu impulsów, jak w „myśleniu" maszynowym. Oznacza to potrzebę zaznajomienia się z prawami, właściwościami czy algorytmami matematycznymi, na podstawie których działają agenci.

Wprowadzanie do ML metod opartych na Human Feedback (HF) daje szanse także do polepszenia rozwiązań skierowanych do prowadzenia analiz pod kątem AML/CFT. Odnieść to można zarówno do „języka mówionego" chatbot (niejednokrotnie zintegrowane z usługami przesyłania wiadomości, aplikacjami na smartfony i stronami internetowymi) czy pisanego. Zarówno w jednym, jak i w drugim obszarze nadal pozostaje wysoka aktywność ze strony IO, takich jak banki.

Dokonywane zapisy treści tej aktywności mogą być następnie przetwarzane na potrzeby identyfikacji tych elementów, które w ramach czynności IO precyzowane są w kierunku oceny klienta/transakcji jako nośnika podejrzeń procederów ML/FT. Zadaniem RLHF mogłoby być wyławianie z wielości sporządzanych dokumentów i prowadzonych rozmów – symptomów podejrzalności ML/FT. Kierunki polityk (taktyki) w tym obszarze z wykorzystaniem HF to analiza zapisów treści oraz języka, za pomocą czego można byłoby identyfikować klienta. Ponadto dokonywanie weryfikacji celu nawiązania relacji z IO, stałego uzupełniania wiedzy w tym zakresie, a także takiego kierowania rozmową, aby można było na bieżąco monitorować klienta w ramach relacji gospodarczych IO – klient. Jednocześnie ta metoda przy odpowiedniej konstrukcji interfejsów pomiędzy dokumentami generowanymi w IO (w obiegu elektronicznym) umożliwiłaby dokonywanie analizy informacji uzyskanych na różnych szczeblach instytucjonalnych oraz w różnych miejscach (geograficznie) aktywności klienta, a także w cyberprzestrzeni. Stosowane w tym zakresie są metody NLP (techniki przetwarzania języka naturalnego), które można łączyć z uczeniem się przez wzmacnianie. Graficzne techniki sztucznej inteligencji, które integrują różne modalności i wykorzystują zależności międzymodalne przez zależności geometryczne, są wykorzystywane do poruszania się po multimodalnych, złożonych danych, tworząc „mapę" połączonych punktów danych. Te zaś ujawniają, w jaki sposób różne typy danych odnoszą się do siebie i wpływają na siebie. RLHF umożliwić może dokonanie diagnostyki zachowań klienta i osób z nim powiązanych, gdy wykryje niezależne zdarzenia w różnych punktach przestępczej aktywności (łączenie śladów). W zakresie budowania środowiska należy mieć na uwadze, że działania sprawców mogą być realizowane w różnych punktach tej samej IO oraz w IO różniących się od siebie. Jest to element taktyki przestępczej, stąd też IO i organy ścigania muszą kreować nowe techniki kontrprzestępcze, do których można zaliczyć także wspieranie się sztuczną inteligencją. Ponadto metody ML/FT zawsze podlegały rozproszeniu działań, tak aby przedstawiciele instytucji nie mogli kojarzyć zdarzeń jako pewnego ciągu logicznego. Od kilku lat jednak dla podwyższenia jakości ochrony IO oraz właściwego dokonywania oceny ryzyka tworzone są wspólne dla rodzajowo tożsamych IO platformy wymiany informacji, które umożliają wzajemne sprawdzanie klientów oraz budowanie tzw. „super SAR" (SuperSARs and information, 2023). Przyjmując rozwiązania, na jakie pozwala RL, wszystkie te obszary można połączyć w środowisko, w którym będzie działał agent (multi-agent) na rzecz rozpoznawania ML/FT (Egli, 2023; Ouyang, Wu, Jiang, 2023).



Rys. 1. Propozycja schematu funkcjonowania RL w instytucji obowiązanej (IO) z wykorzystaniem multi-agenta i HF
Źródło: opracowanie własne

Raport z sierpnia 2023 r., opracowany przez Europejski Urząd Nadzoru Bankowego (ang. The European Banking Authority, EBA), wskazał, że uczenie maszynowe w zakresie podstawowego modelowania (tj. różnicowanie ryzyka i kwantyfikacja ryzyka), mające charakter wewnętrzny, instytucje obowiązane wykorzystywały głównie do różnicowania ryzyka w zakresie ML/FT (Raport EBA, 2023). W przypadku danych wejściowych techniki ML mogą być przydatne w celu wybrania odpowiednich zmiennych do modelu przez, gdy na przykład nastąpi przekształcanie nowych danych (metody eksploracji tekstu), wystąpi potrzeba uzupełniania brakujących danych oraz ich przetwarzanie oraz wykorzystywanie danych nieustrukturyzowanych. Należy także zwrócić uwagę na to, że ML powoduje zwiększenie wydajności modeli oceny ryzyka. Uczenie maszynowe można także wykorzystać w celu poprawy wydajności modeli pretendentów i jako analizę pomocniczą w odniesieniu do alternatywnych założeń lub podejść, a także na potrzeby walidacji danych.

W ramach zaleceń EBA dla instytucji wskazano, że ML może być wykorzystywane do różnych celów i na różnych poziomach, na przykład przygotowywania danych, różnicowania ryzyka, kwantyfikacji ryzyka i celów wewnętrznej walidacji. Jeśli instytucje chcą stosować modele uczenia maszynowego do celów kapitału regulacyjnego, to wszyscy właściwi interesariusze powinni posiadać odpowiedni poziom wiedzy na temat funkcjonowania modelu. Ponadto zalecono, aby instytucje znalazły odpowiednią równowagę pomiędzy wynikami modelu a wyjaśnieniami wyników. Wyższy poziom złożoności może prowadzić do lepszego modelu wydajności, ale kosztem mniejszej wyjaśnialności i zrozumienia modelu funkcjonowania. Dlatego instytucjom zaleca się unikanie niepotrzebnej złożoności w procesie podejścia modelowego, jeśli nie jest to uzasadnione znaczącą poprawą zdolności predykcyjnych. Dotyczy to przede wszystkim niestosowania nadmiernej liczby czynników wyjaśniających, niekorzystanie z nieustrukturyzowanych danych, jeżeli można korzystać z bardziej konwencjonalnych informacji, oraz unikanie zbyt złożonych modeli, jeżeli prostsze dają także pozytywne wyniki. Należy zauważyć, że stanowisko zaprezentowane przez EBA wobec tego, jak wiele dokonały ośrodki naukowe na rzecz rozwoju sztucznej inteligencji, w tym w zakresie RL, jest podejściem ostrożnościowym. Wydaje się, że wynika ono z różnego przygotowania poszczególnych IO do podejmowania współdziałania z modelami uczenia maszynowego, a także z założenia, że jeśli dana instytucja decyduje się na użycie modelu ML, to przede wszystkim musi znać jego specyfikę i posiadać wiedzę i interpretację uzyskanych za pomocą takiego modelu wyników. Oznacza to, że IO nie może bezkrytycznie podchodzić do uzyskanych w wyniku ML danych i powinna wiedzieć, jak je użyć na potrzeby procesu AML/CFT.

Podsumowanie

System przeciwdziałania ML/FT stanowi skomplikowane środowisko czynników, które nie pozostają jednocześnie obojętne wobec zmieniającej się rzeczywistości. Ponadto środowisko to generuje codziennie, zwłaszcza wobec niektórych IO, wielość danych, które należy poddać analizie na potrzeby wyłowienia z nich zachowań umożliwiających ocenę jako podejrzenie procederu ML/FT. W konsekwencji od lat prowadzone są poszukiwania wśród najnowszych technik, przede wszystkim wśród uczenia maszynowego, metod wsparcia dla czynnika ludzkiego (analitycznego), którym można byłoby wesprzeć dotychczasowe techniki przeciwdziałania. Jest to między innymi wynikiem małej efektywności dotychczasowych metod jakościowych/ilościowych w zakresie generowania raportów o podejrzanych transakcjach i kierowania ich do jednostek analityki finansowej. Dodatkowo rozwija się także technicznie platforma oferowania różnych produktów i usług online z wykorzystaniem nowoczesnych technik pomnażania aktywów. Wśród samych metod uczenia maszynowego uzyskuje się postęp, również w kierunku wsparcia rozwiązań analitycznych. Uczenie maszynowe przez wzmocnienie jest dobrym przykładem różnorodności wsparcia oferowanego dla komórek AML/CFT funkcjonujących w IO. Co istotne, już nie tylko metody te działają na zasadzie wygenerowania powtarzającej się informacji (inteligentne kopiowanie), lecz także na potrzeby alternatywnego „myślenia" ze wskazanym kierunkiem (polityką) uzyskiwania informacji. Uwzględniając wspomnianą różnorodność środowiska AML/CFT w IO, ML pozwala zarówno postępować jedynie wśród rozwiązań technicznych, jak i z czynnym udziałem czynnika ludzkiego, jeszcze w fazie predecyzyjnej. Należy jednak zauważyć, że oprócz zautomatyzowanych rozwiązań opartych na regułach (transakcje ponadprogowe) w pozostałych przypadkach oprócz „podpowiedzi" ze strony agenta czynnik ludzki jako decyzyjny w relacji z JAF staje się niezbędny. Wydaje się, że stan N (związany z powstaniem podejrzalności ML/FT), któremu należałoby zapobiec, co w dużej mierze staje się rozpoznawalne z opóźnieniem t, jeżeli nie będziemy go obserwowali w czasie N(t), a jeżeli tak, to raczej IO zależałoby na niedoprowadzeniu do zdarzenia Z w czasie T (N_t) niż godzenie się na jego dokonanie. Tym samym przedstawione rozwiązania raczej zmierzają do uzyskania stanu przekonywalności, a więc sprowadzenia możliwego zachowania do prawdopodobieństwa jego powstania w przewidywalnej formie (prowadzą do całkowitego wyeliminowania stanu niepewności). Przy czym w większości przypadków mimo ich przewidywalności nie osiągniemy całkowitej pewności, chyba że działanie sprawcy sprowadzone zostanie jedynie do odwzorowania formy (np. przygotowania lub dokonania czynu) i w pełni przewidywalnym punkcie czasowym i miejsca wystąpienia takiego zachowania (uzyskanie prawdopodobieństwa wystąpienia zdarzenia uzyskanego na podstawie posiadanych danych). Do ustalenia prawdopodobieństwa zdarzenia przyczynić się zaś powinna także ocena ryzyka wprowadzona jako jeden

z obowiązków wykonawczych w systemie AML/CFT dla IO, zwłaszcza w przypadku oceny ryzyka indywidualnego klienta. Przyspieszenia procesów analitycznych można natomiast dokonać przez MARL, czyli współdziałanie poszczególnych agentów. Platformą mogą być zarówno te same dane (pochodzące z tego samego środowiska) lub różne dane (uplasowane w niezależnych środowiskach). Spoiwem dla działań analitycznych będzie więc informacja lub kooperacja agentów. Przyszłościowo można wskazać, czy nie podążyć innym tropem niż dotychczasowe rozwiązania w zakresie AML/CFT (jako kontrreakcja uwzględniająca ryzyko i stosująca środki bezpieczeństwa finansowego) i oprzeć je np. na Chat GPT (należy mieć na uwadze to, że stale poszerza się zakres usług finansowych świadczonych przez instytucje za pośrednictwem chatbotów) jako nowej metodzie oferowania produktów i usług klientom z wbudowanymi środkami bezpieczeństwa, tym samym umożliwiając takie sterowanie produktami i usługami, aby nie dopuścić do ich angażowania się w proceder prania pieniędzy czy finansowania terroryzmu. Sterowanie byłoby oparte na kierowaniu procesami biznesowymi w bankowości i usługach finansowych jedynie w razie pozytywnego scenariusza, tj. niewchodzenia w konflikt z prawem, z zachowaniem prawa do innowacyjności obydwu stron.

BIBLIOGRAFIA

[1] Abramson, J., Ahuja, A., Carnevale, F., Georgiev, P., 2022. *Improving Multimodal Interactive Agents with Reinforcement Learning from Human Feedback*, https://arxiv.org/pdf/2211.11602.pdf (dostęp: 22.11.2023).

[2] Bartuś, T., 2013. Zastosowanie inteligentnych agentów w administracji publicznej, Wydział Ekonomii Uniwersytet Ekonomiczny w Katowicach, *Roczniki Kolegium Analiz Ekonomicznych*, nr 29.

[3] Dhaduk, H., 2023. *A Complete Guide to Fine Tuning Large Language Models. Simform – Product Engineering Company*, https://www.simform.com/blog/completeguide-finetuning-llm/ (dostęp: 20.11.2023).

[4] Eastnets, 2023. *Is open-source AI a good or bad thing for the finance sector?*, https://www.eastnets.com/newsroom/is-open-source-ai-a-good-or-bad-thing-for-the-finance-sector (dostęp: 24.11.2023).

[5] Egli, A., 2023. ChatGPT, GPT-4, and Other Large Language Models: The Next Revolution for Clinical Microbiology?, *Clinical Infectious Diseases*, vol. 77, nr 9.

[6] Frąckiewicz, M., 2023. *Przeciwstawne uczenie maszynowe*, https://ts2.space/pl/przeciwstawne-uczenie-maszynowe/#gsc.tab=0 (dostęp: 26.11.2023).

[7] Guoxinag, T., Jieyu, S., 2023. Financial transaction fraud detector based on imbalance learning and graph neural network, *Applied Soft Computing*, vol. 149, Part A.

[8] Gupta, J.K., Egorov, M., Kochenderfel, M., 2017. *Cooperative Multi-Agent Control Using Deep Reinforcement Learning*, pkt 4.1, https://ala2017.cs.universityofgalway.ie/papers/ALA2017_Gupta.pdf (dostęp: 26.11.2023).

[9] Wikipedia, 2013. *Mapa wektorowa*, https://pl.wikipedia.org/wiki/Mapa_wektorowa (dostęp: 26.11.2023).

[10] Mehta, K., Mahajan, A., Kumar, P., 2023. *marl-jax: Multi-Agent Reinforcement Leaning Framework*, https://arxiv.org/pdf/2303.13808.pdf (dostęp: 28.11.2023).

[11] Muller, A.C., Guido, S., 2023. *Machine learning, Python i data science*, Gliwice: Wydawnictwo Helion.

[12] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C.L., Mishkin, P., Zhang, Ch., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., Lowe, R., 2023. *Training language models to follow instructions with human feedback*, https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf (dostęp: 28.11.2023).

[13] Patrizo, A., 2023. *Reinforcement learning from human feedback (RLHF)*, https://www.techtarget.com/whatis/definition/reinforcement-learning-from-human-feedback-RLHF, (dostęp: 28.11.2023).

[14] Qiu, W., Ma, X., An, B., Obraztsova, S., Yan, S.H., Xu, Z., 2023, RPM: *Generalizable Multi-Agent Policies For Multi-Agent Reinforcement Learning*, https://arxiv.org/pdf/2210.09646.pdf (dostęp: 28.11.2023).

[15] Raport EBA, 2023. *Machine Learning for IRB Models. Follow-Up Report From The Consultation On The Discussion Paper On Machine Learning for IRB Models,* Eba/Rep/2023/28, August 2023, https://www.eba.europa.eu/sites/default/documents/files/document_library/Publications/Reports/2023/1061483/Follow-up%20report%20on%20machine%20learning%20for%20IRB%20models.pdf (dostęp: 25.11.2023).

[16] Standen, M., Kim, J., Szabo, C., 2023. *SoK: Adversarial Machine Learning Attacks and Defences in Multi-Agent Reinforcement Learning*, https://arxiv.org/abs/2301.04299 (dostęp: 25.11.2023).

[17] SuperSARs and information, 2023. *SuperSARs and information sharing in the regulated sector*, https://www.comsuregroup.com/news/supersars-and-information-sharing-in-the-regulated-sector/ (dostęp: 27.11.2023).

[18] Tong, G., Shen, J., 2023. Financial transaction fraud detector based on imbalance learning and graph neural network, *Applied Soft Computing*, vol. 149, Part A.

[19] Weyns, D., 2010. *Architecture-Based Design of Multi-Agent Systems*, Berlin–Heidelberg: Springer-Verlag.

[20] Zhang, K., Yang, Z., Başar, T., 2021. *Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms*, https://arxiv.org/pdf/1911.10635.pdf (dostęp: 22.11.2023).

[21] Zhao, B., Jin, W., Chen, Z., Guo, Y., 2023. *A semi-independent policies training method with shared representation for heterogeneous multi-agents reinforcement learning*, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2023.1201370/full (dostęp: 22.11.2023).