Justyna Wójtowicz-Rutkowska*, Magdalena Litwinowicz, Wiesław Capała, Paweł Łysik, Michał Jeromkin

Sieć Badawcza Łukasiewicz – Instytut Chemii Przemysłowej imienia Profesora Ignacego Mościckiego, Warszawa

Pitfalls and risks of implementing artificial intelligence in the chemical industry

Pułapki i ryzyka wdrażania sztucznej inteligencji w przemyśle chemicznym

OPEN ACCESS **DOI:** *10.15199/62.2026.5.6*

Technol. and organizational risks associated with use of artificial intelligence were analyzed. Data qual. limitations, restricted model applicability, performance degradation over time, and implications for process safety were taken into consideration. Particular attention was given to overreliance on algorithmic recommendations and the inherent limitations of models trained on historical data.

Systemy sztucznej inteligencji są coraz częściej integrowane z instalacjami chemicznymi jako narzędzia wspierające predykcyjne utrzymanie ruchu, monitorowanie procesu oraz optymalizację pracy. Ich zastosowanie wpływa na sposób podejmowania decyzji operacyjnych oraz może modyfikować podział odpowiedzialności w organizacji. Przeanalizowano technologiczne i organizacyjne ryzyka związane z wykorzystaniem AI, w tym ograniczenia wynikające z jakości danych, zakresu stosowalności modeli oraz utraty trafności w czasie, a także konsekwencje dla bezpieczeństwa procesu. Zwrócono uwagę na zjawisko nadmiernego polegania na rekomendacjach algorytmów oraz na ograniczenia modeli trenowanych na danych historycznych. W instalacjach przemysłowych odpowiedzialność za decyzje techniczne pozostaje po stronie zespołu prowadzącego proces. Narzędzia AI stosowane w instalacjach chemicznych nie stanowią warstwy bezpieczeństwa w rozumieniu norm bezpieczeństwa funkcjonalnego, dlatego wymagają formalnego nadzoru w całym cyklu życia oraz włączenia do procedur zarządzania zmianą MOC (*management of change*).

***Keywords**: artificial intelligence, chemical plants, process safety, data quality, operational decision-making, human oversight*

Słowa kluczowe: sztuczna inteligencja, instalacje chemiczne, bezpieczeństwo procesowe, jakość danych, decyzje operacyjne, zarządzanie ryzykiem

Eksploatacja instalacji chemicznych odbywa się w warunkach ograniczonego marginesu operacyjnego, w których parametry procesu muszą jednocześnie spełniać wymagania technologiczne oraz kryteria bezpieczeństwa. Nawet niewielkie odchylenia parametrów lub niewłaściwe decyzje operacyjne mogą prowadzić do pogorszenia stabilności pracy instalacji oraz stopniowego zmniejszenia dostępnej rezerwy bezpieczeństwa. Parametry mogą mieścić się w dopuszczalnym zakresie, mimo że instalacja nie pracuje już w bezpiecznym komforcie operacyjnym. W takich sytuacjach operator może obserwować wolniejszą reakcję układu regulacji na zmianę nastawy, niestabilność

Mgr inż. Justyna WÓJTOWICZ-RUTKOWSKA (ORCID: 0000-0002-0276-1108) w roku 1998 ukończyła studia na Wydziale Inżynierii i Technologii Chemicznej Politechniki Krakowskiej. Zawodowo jest związana z Sieć Badawcza Łukasiewicz – Instytutem Chemii Przemysłowej imienia Profesora Ignacego Mościckiego w Warszawie, gdzie obecnie pracuje na stanowisku starszego specjalisty pionu badawczego w Sekcji Procesów Katalitycznych. Specjalność – inżynieria chemiczna.

Dr inż. Magdalena LITWINOWICZ (ORCID: 0000-0001-9058-4764) w roku 2002 ukończyła studia na Wydziale Chemicznym Politechniki Gdańskiej, a w 2018 r. uzyskała stopień doktora nauk technicznych na Wydziale Chemicznym Politechniki Śląskiej w Gliwicach. Od 2004 r. jest związana z Sieć Badawcza Łukasiewicz – Instytutem Chemii Przemysłowej imienia Profesora Ignacego Mościckiego w Warszawie, gdzie obecnie pracuje na stanowisku lidera obszaru w Pionie Badawczym i kierownika Sekcji Procesów Katalitycznych. Specjalność – technologia chemiczna i kataliza.

*** Adres do korespondencji:**
Sieć Badawcza Łukasiewicz – Instytut Chemii Przemysłowej imienia Profesora Ignacego Mościckiego, ul. Rydygiera 8, 01-793 Warszawa, tel.: +48 517-883-159,
e-mail: justyna.wojtowicz-rutkowska@ichp.lukasiewicz.gov.pl

odpowiedzi aparatury lub inne symptomy pogorszenia stabilności pracy instalacji.

W ostatnich latach w przemyśle chemicznym coraz częściej wdrażane są systemy sztucznej inteligencji wspierające analizę danych procesowych, predykcyjne utrzymanie ruchu oraz optymalizację pracy instalacji. Systemy te nie stanowią elementów układów regulacji ani warstw zabezpieczeń procesowych[1, 2]. Pełnią funkcję narzędzi analitycznych generujących rekomendacje na podstawie danych procesowych. Najczęściej są to modele oparte na analizie danych historycznych, wykorzystywane do predykcji stanów technicznych urządzeń, identyfikacji odchyleń procesowych oraz wspomagania optymalizacji parametrów pracy.

Modele uczenia maszynowego wykorzystują sygnały procesowe i dane historyczne dostępne w systemach pomiarowych i systemach archiwizacji danych. Klasyczne systemy sterowania i zabezpieczeń procesowych projektowane są w sposób deterministyczny i poddawane formalnej analizie scenariuszy awaryjnych. Modele oparte na uczeniu maszynowym mają natomiast charakter probabilistyczny i zależą od struktury danych, na których zostały wytrenowane. Jeżeli rzeczywiste warunki pracy instalacji odbiegają od zakresu reprezentowanego w danych treningowych, model może generować rekomendacje nieadekwatne do aktualnego stanu procesu.

Wprowadzenie takich narzędzi do środowiska operacyjnego instalacji chemicznej zmienia charakter niepewności związanej z podejmowaniem decyzji operacyjnych. Zamiast niepewności analizowanej na etapie projektowym pojawia się niepewność estymowana statystycznie, zależna od jakości danych oraz zakresu walidacji modelu. Architekturę decyzyjną instalacji chemicznej z uwzględnieniem warstwy analitycznej AI przedstawiono schematycznie na rys. 1.

Znaczenie tych zagadnień znajduje również odzwierciedlenie w aktualnych regulacjach europejskich dotyczących systemów sztucznej inteligencji (AI Act)[3], które wprowadzają klasyfikację rozwiązań wg poziomu ryzyka oraz wymagają zapewnienia nadzoru nad ich stosowaniem w całym cyklu życia. Z punktu widzenia instalacji chemicznych szczególne znaczenie mają zastosowania mogące wpływać na sposób prowadzenia procesu oraz bezpieczeństwo pracy instalacji. W takich przypadkach systemy AI mogą podlegać wymaganiom właściwym dla systemów wysokiego ryzyka, obejmującym m.in. konieczność zapewnienia jakości danych, przejrzystości działania oraz nadzoru człowieka. W praktyce przemysłowej oznacza to konieczność jednoznacznego określenia zakresu zastosowania modelu oraz warunków jego wykorzystania operacyjnego. Dotyczy to w szczególności sytuacji, w których rekomendacje algorytmu wpływają na sposób prowadzenia procesu lub interpretację stanu instalacji. W takich przypadkach konieczne jest również określenie zasad postępowania poza zakresem walidacji modelu oraz warunków jego ponownej oceny po zmianach technologicznych, jakości surowca lub konfiguracji aparatury.

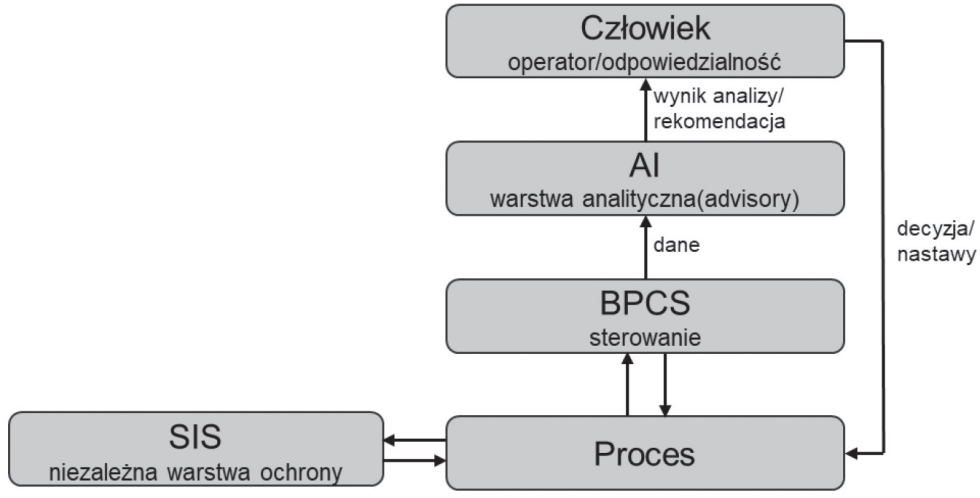
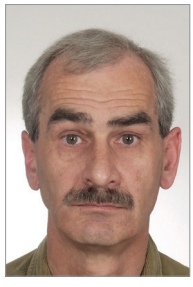


Fig. 1. Decision layers in a chemical plant including the AI system (author's own elaboration based on[1,2])

Rys. 1. Warstwy decyzyjne w instalacji chemicznej z uwzględnieniem systemu AI (opracowanie własne na podstawie[1, 2])



Mgr inż. Wiesław CAPAŁA (ORCID: 0000-0002-7701-8853) w roku 1986 ukończył studia na Politechnice Warszawskiej. Od 1988 r. jest zatrudniony w Sieć Badawcza Łukasiewicz – Instytucie Chemii Przemysłowej imienia Profesora Ignacego Mościckiego w Warszawie, obecnie na stanowisku głównego specjalisty pionu wsparcia w Sekcji Procesów Katalitycznych. Specjalność – inżynieria chemiczna i procesowa.

Mgr inż. Paweł ŁYSIK (ORCID: 0000-0002-6667-2622) w roku 2000 ukończył studia na Wydziale Inżynierii Chemicznej i Procesowej Politechniki Warszawskiej. Od tego czasu jest zatrudniony w Sieć Badawcza Łukasiewicz – Instytucie Chemii Przemysłowej imienia Profesora Ignacego Mościckiego w Warszawie, obecnie na stanowisku starszego specjalisty. Specjalność – inżynieria chemiczna i procesowa.

Kiedy rekomendacja staje się decyzją

Na początkowym etapie wdrożenia system AI pełni funkcję pomocniczą: analizuje dane, wskazuje odchylenia i sugeruje korekty parametrów. Dopóki jego rola jest jednoznacznie określona, proces decyzyjny pozostaje przejrzysty, a rekomendacja stanowi jedno ze źródeł informacji. Ryzyko wzrasta, gdy rekomendacja modelu staje się domyślną podstawą decyzji operacyjnej, a jej niezależna weryfikacja jest osłabiana. Taka zmiana zwykle nie następuje jednorazowo. Zaufanie do systemu narasta stopniowo, zwłaszcza gdy przez dłuższy czas rekomendacje wydają się trafne. W konsekwencji potrzeba ich niezależnej weryfikacji może słabnąć. Mechanizm ten odpowiada opisywanemu w literaturze zjawisku *automation bias*, polegającemu na nadmiernym zaufaniu do systemów wspomagania decyzji oraz ograniczeniu niezależnej oceny ich wskazań[4, 5]. W praktyce analiza sytuacji operacyjnej bywa inicjowana wynikiem modelu, a dalsza interpretacja danych sprowadza się do weryfikacji zgodności z tą rekomendacją. Model może sugerować niewielką korektę parametru mieszczącą się w dopuszczalnym zakresie. Formalnie działanie pozostaje zgodne z procedurą, jednak taka zmiana może zmniejszać rezerwę operacyjną i przesuwać proces w rzadko eksploatowany obszar pracy instalacji. Ponieważ model opiera się na danych historycznych, nie uwzględnia wszystkich bieżących ograniczeń technicznych, w tym zmian charakterystyki aparatury i dynamiki procesu. Wraz ze wzrostem wpływu systemu zmienia się również sposób uzasadniania decyzji operacyjnych: coraz częściej odwołują się one do wyniku modelu. Jeżeli działanie modelu nie jest przejrzyste, jego wynik bywa trudniejszy do zakwestionowania. W dokumentacji łatwo wskazać zgodność z rekomendacją systemu, natomiast trudniej jednoznacznie określić, kto faktycznie podjął decyzję[4, 5]. Formalna odpowiedzialność pozostaje po stronie zespołu operacyjnego[6–8], jednak realny wpływ systemu na przebieg procesu decyzyjnego może być istotny. W instalacjach o niewielkim marginesie bezpieczeństwa niejasne określenie roli systemu AI stanowi dodatkowe ryzyko operacyjne.

Ryzyka technologiczne

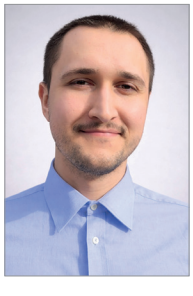
Jakość danych i zakres stosowalności modeli

Poprawne działanie modelu nie jest równoznaczne z poprawnością merytoryczną jego wyniku. Błędy predykcji mogą występować w szczególności wtedy, gdy model jest stosowany poza zakresem warunków, w których został wytrenowany i zwalidowany. Problem ograniczonej zdolności generalizacji modeli oraz ich stosowalności poza obszarem danych treningowych jest szeroko omawiany w literaturze dotyczącej zastosowań uczenia maszynowego w inżynierii procesowej[9, 10].

W inżynierii procesowej modele te uczą się na danych pochodzących z określonego okresu pracy instalacji[9, 10]. Obejmują jedynie te stany, które rzeczywiście wystąpiły przy konkretnych surowcach, konfiguracji aparatury oraz sposobie prowadzenia procesu. Zależności identyfikowane przez model odzwierciedlają relacje obecne w zbiorze danych, a nie pełną przestrzeń możliwych stanów procesu. Podkreślają to również analizy wdrożeń przemysłowych[9, 10]. W praktyce instalacja bywa eksploatowana w szerszym zakresie, niż obejmuje to zbiór danych. Przekroczenie tego zakresu nie zatrzymuje obliczeń ani nie musi być sygnalizowane przez system. W literaturze uczenia maszynowego problem ten opisuje się jako zastosowanie modelu do danych spoza rozkładu treningowego OOD (*out-of-distribution*), bez gwarancji poprawności predykcji[11]. W takiej sytuacji system może nadal generować wynik, nie sygnalizując, że dane znajdują się poza zakresem jego walidacji. Brak komunikatu o błędzie nie stanowi potwierdzenia poprawności ani bezpieczeństwa rekomendacji. W raportach bezpieczeństwa procesowego wielokrotnie wskazywano, że nieprawidłowy obraz procesu może utrzymywać się mimo formalnie poprawnych wskazań systemów pomiarowych. Analizy poważnych zdarzeń procesowych pokazują, że brak jednoznacznego sygnału ostrzegawczego nie jest równoznaczny z bezpiecznym stanem instalacji, co podkreślono m.in. w analizie zdarzenia w rafinerii Texas City[12]. Choć zdarzenie to nie dotyczyło systemów AI, pokazuje ono ogólny mechanizm interpretacji sygnałów procesowych w warunkach niejednoznacznej informacji. Mechanizm ten ma znaczenie również w kontekście modeli uczonych na danych historycznych. Algorytm może generować wynik poprawny obliczeniowo, lecz nieadekwatny do aktualnego reżimu pracy. Jeżeli system nie sygnalizuje wyjścia poza zakres walidacji, użytkownik otrzymuje informację pozornie kompletną. W środowisku wysokiego ryzyka brak jednoznacznego sygnału nie może być traktowany jako brak zagrożenia[6].

Utrzymanie trafności modelu w czasie

Trafność modelu nie jest wartością stałą. Parametry procesu mogą zmieniać się w czasie, często szybciej niż zakładano na etapie przygotowania modelu, np. wskutek zmian surowców, modyfikacji aparatury lub zmiennego obciążenia. System może pozostawać operacyjnie dostępny, lecz jego użyteczność stopniowo się pogarsza. W literaturze zjawisko to określane jest jako *data drift* oraz *concept drift*[11]. W praktyce oznacza to zmianę charakterystyki danych wejściowych lub zmianę zależności między zmiennymi



Mgr inż. Michał JEROMKIN (ORCID: 0000-0002-7924-9786) w roku 2017 ukończył studia na Wydziale Inżynierii Chemicznej i Procesowej Politechniki Warszawskiej. Od 2018 r. pracuje w Sieć Badawcza Łukasiewicz – Instytucie Chemii Przemysłowej imienia Profesora Ignacego Mościckiego w Warszawie, obecnie jako starszy specjalista w sekcji Procesów Katalitycznych. Specjalność – inżynieria procesów ochrony środowiska.

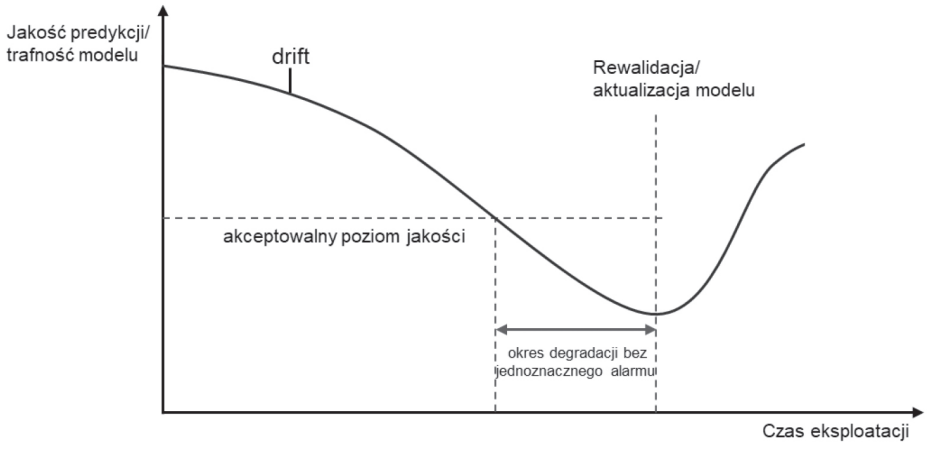


Fig. 2. Conceptual illustration of model performance degradation over time and the role of revalidation (author's own elaboration based on[9, 11–13])

Rys. 2. Koncepcyjna ilustracja degradacji trafności modelu w czasie oraz roli rewalidacji (opracowanie

procesowymi. W warunkach przemysłowych oba mechanizmy często współwystępują; niekiedy wystarczy korekta kalibracji czujnika lub zmiana jakości partii surowca[9, 10]. Model zwykle nie sygnalizuje wprost spadku trafności. Może nadal generować wynik, a sygnały procesowe prezentowane na wykresach pozostają pozornie poprawne. Taka sytuacja utrudnia identyfikację narastającej rozbieżności między rzeczywistością a predykcją. W literaturze dotyczącej zastosowań modeli w środowisku przemysłowym podkreśla się konieczność monitorowania jakości predykcji oraz okresowej weryfikacji i rewalidacji modelu w zmieniających się warunkach eksploatacyjnych[9–11]. Degradacja jakości ma zwykle charakter powolny i nie wskazuje jednego momentu, w którym można jednoznacznie stwierdzić utratę wiarygodności modelu. W efekcie łatwo przyjąć, że skoro model działał poprawnie przez dłuższy czas, to pozostaje wiarygodny, mimo że proces mógł już wejść w odmienny reżim pracy. W konsekwencji decyzje mogą być podejmowane na podstawie narzędzia opisującego warunki sprzed kilku miesięcy[11]. Zjawisko to przedstawiono schematycznie na rys. 2.

Wdrożenie modelu nie zamyka prac nad jego utrzymaniem. Trafność predykcji wymaga weryfikacji w czasie, analogicznie do kontroli stanu aparatury lub stabilności procesu[9, 10]. Model można traktować jako element infrastruktury analitycznej, który ulega „starzeniu" w zmieniających się warunkach eksploatacyjnych. W literaturze opisano analizę modeli predykcyjnych zużycia energii w maszynach przemysłowych opartą na danych operacyjnych z kilku obiektów produkcyjnych. Model statyczny, trenowany na danych historycznych, osiągał w warunkach walidacyjnych średni błąd MAPE rzędu 7–9%, natomiast w kolejnych okresach eksploatacji obserwowano odcinki pogorszenia jakości predykcji wraz ze zmianą obciążenia i charakterystyki pracy urządzeń. Zastosowanie mechanizmu wykrywania *concept drift* oraz adaptacyjnego dostrajania modelu pozwoliło obniżyć średni błąd do 4–5%[13]. Przykład ten wskazuje, że bez monitorowania jakości predykcji i detekcji zmiany reżimu pracy model może zachować poprawność obliczeniową przy jednoczesnej utracie adekwatności do warunków rzeczywistych.

Granice integracji z systemem sterowania i bezpieczeństwa

Systemy AI najczęściej wprowadzane są do instalacji jako narzędzia analityczne. Z czasem ich wskazania mogą jednak wpływać na nastawy regulatorów lub sposób prowadzenia węzła technologicznego, a granica między analizą a oddziaływaniem na proces przestaje być jednoznaczna. Instalacja chemiczna wykorzystuje podstawowy system sterowania procesem BPCS (*basic process control system*) oraz system instrumentowanych funkcji bezpieczeństwa SIS (*safety instrumented system*). Warstwa SIS jest projektowana zgodnie z wymaganiami bezpieczeństwa funkcjonalnego, m.in. wg normy IEC oraz wytycznych dotyczących bezpiecznej automatyzacji procesów chemicznych[7, 9]. Jej zadaniem jest zapewnienie przewidywalnej reakcji na zdefiniowane scenariusze oraz określonego poziomu zmniejszania ryzyka, a jej elementy podlegają formalnej analizie i walidacji[7].

Modele oparte na uczeniu maszynowym powstają w odmiennym reżimie projektowym. Systemy zabezpieczeń procesowych projektowane są deterministycznie, z jasno zdefiniowanymi funkcjami bezpieczeństwa i formalną weryfikacją poziomu zmniejszania ryzyka. Modele uczenia maszynowego mają natomiast charakter statystyczny i probabilistyczny, a ich działanie zależy od rozkładu danych historycznych oraz przyjętych założeń treningowych[9, 10]. Jeżeli wpływ takiego modelu na sposób prowadzenia procesu rośnie, jego wdrożenie oraz kolejne etapy integracji powinny podlegać formalnym procedurom zarządzania zmianą (MOC) oraz ocenie wpływu na ryzyko procesowe[8]. System AI nie jest tworzony jako element bezpieczeństwa funkcjonalnego i nie przechodzi procedur wymaganych dla warstwy SIS[6]. Mimo to może pośrednio wpływać na parametry pracy instalacji, a w konsekwencji na poziom ryzyka. Jeżeli algorytm zaczyna realnie oddziaływać na proces, jego wdrożenie należy traktować jak zmianę technologiczną i objąć formalną oceną wpływu na ryzyko, zgodnie z zasadami bezpieczeństwa procesowego[6–8]. W przeciwnym razie powstaje element wpływający na proces, który nie został oceniony wg zasad właściwych dla systemów bezpieczeństwa. W praktyce granica ta przesuwa się stopniowo: od rekomendacji, przez automatyczne przekazywanie sugestii operatorowi, aż po częściową automatyzację decyzji. Zastosowania uczenia maszynowego obejmują również predykcję i klasyfikację alarmów

procesowych w rzeczywistych obiektach przemysłowych. W opisanym w literaturze przypadku z zakładu gazowego, modele nadzorowane wykorzystano do ograniczenia liczby nieistotnych sygnałów oraz wsparcia zarządzania alarmami i interpretacji stanów operacyjnych[14]. Takie rozwiązania, choć formalnie mają charakter analityczny, wpływają na sposób oceny sytuacji procesowej i mogą przesuwać granicę między analizą a oddziaływaniem na prowadzenie instalacji. Każdy z tych etapów może wydawać się niewielki, jednak łącznie prowadzą do sytuacji, w której system nieprojektowany jako warstwa ochronna zaczyna istotnie wpływać na poziom ryzyka. System AI może wspierać analizę danych i podejmowanie decyzji, lecz nie stanowi zabezpieczenia procesowego w rozumieniu norm bezpieczeństwa funkcjonalnego[7]. Podobne problemy integracji nowych narzędzi cyfrowych z systemami bezpieczeństwa procesowego wskazywane są również w literaturze dotyczącej bezpieczeństwa procesowego[7–9].

Ograniczenia infrastrukturalne i integracyjne

Model pracuje na danych, które instalacja udostępnia. Jeżeli system pomiarowy jest ograniczony lub niespójny, obraz procesu pozostaje niepełny, niezależnie od jakości zastosowanego algorytmu. W wielu obiektach systemy archiwizacji danych projektowano z myślą o bieżącej eksploatacji, a nie o budowie narzędzi predykcyjnych. Historia zmian konfiguracji, modernizacji i modyfikacji aparatury nie zawsze jest jednoznacznie powiązana z danymi procesowymi. W starszych instalacjach część działań korygujących wykonywana jest ręcznie lub lokalnie, bez pełnego odzwierciedlenia w systemie sterowania. Stabilność pracy bywa wówczas efektem doświadczenia zespołu, a nie wyłącznie działania automatyki, czego dane historyczne nie oddają w pełni. W takich warunkach wdrożenia AI mają często charakter punktowy i obejmują wybrane fragmenty instalacji. Integracja nowych narzędzi cyfrowych z istniejącą infrastrukturą napotyka ograniczenia związane ze stanem systemów pomiarowych, archiwizacji danych oraz historią modyfikacji instalacji. Problemy te wskazywano w opracowaniach dotyczących modernizacji istniejących obiektów przemysłowych[15]. W literaturze dotyczącej transformacji cyfrowej obiektów typu *brown field* (obiektów istniejących, modernizowanych) podkreśla się, że skuteczność wdrożenia zależy w równym stopniu od jakości danych i spójności infrastruktury, jak i od samego algorytmu[16]. W praktyce oznacza to, że trudności wdrożeniowe częściej wynikają z niespójności infrastruktury danych oraz integracji systemów niż z ograniczeń samego algorytmu. O powodzeniu wdrożenia decyduje nie tylko jakość modelu, lecz także stan techniczny obiektu oraz sposób zarządzania danymi. Ocena modelu nie powinna ograniczać się wyłącznie do jakości predykcji, lecz powinna uwzględniać warunki pozyskiwania danych oraz ich reprezentatywność względem aktualnego stanu instalacji. Dotyczy to zwłaszcza obiektów o długiej historii eksploatacji, w których zmiany konfiguracji aparatury lub modernizacje nie są w pełni odzwierciedlone w danych historycznych.

W obiektach o długiej historii eksploatacji każda zmiana technologiczna powinna być oceniana w kontekście istniejącej infrastruktury oraz praktyk prowadzenia procesu. Ważnym elementem jest również przygotowanie organizacyjne, obejmujące rozwój kompetencji personelu oraz zarządzanie zmianą podczas wdrażania systemów AI. W kontekście eksploatacji instalacji przemysłowych szczególnego znaczenia nabiera zapewnienie jakości danych wykorzystywanych przez modele. Zmiany charakterystyki danych procesowych, wynikające np. z modyfikacji warunków pracy instalacji lub parametrów pomiarowych, mogą prowadzić do pogorszenia jakości predykcji i zwiększenia ryzyka błędnych decyzji operacyjnych. Ryzyka związane z zastosowaniem systemów AI mają charakter wielowymiarowy i obejmują nie tylko aspekty technologiczne, lecz także wpływ na stabilność operacyjną instalacji oraz bezpieczeństwo prowadzenia procesów. Ocena tych ryzyk wymaga uwzględnienia zarówno ograniczeń technologicznych, jak i sposobu prowadzenia procesu oraz organizacji pracy.

W przedstawionym kontekście jakości i spójności danych szczególnego znaczenia nabierają również zagrożenia związane z cyberbezpieczeństwem systemów AI. Modele uczenia maszynowego są bezpośrednio zależne od danych treningowych oraz danych dostarczanych w trakcie eksploatacji. Manipulacja danymi treningowymi może utrwalać w modelu zależności, które nie odzwierciedlają rzeczywistych mechanizmów procesu[17]. Zakłócenia bieżących danych procesowych mogą natomiast wpływać na rekomendacje systemu, bez jednoznacznych sygnałów błędu. Ważne jest też bezpieczeństwo infrastruktury danych oraz samego modelu. W warunkach przemysłowych możliwa jest ingerencja zarówno w dane, jak i w sposób działania modelu, co może prowadzić do podejmowania decyzji operacyjnych na podstawie nieprawidłowego obrazu procesu.

W takich przypadkach kluczowe staje się zapewnienie integralności danych oraz możliwość odtworzenia podstaw generowania rekomendacji modelu. Obejmuje to kontrolę źródeł danych oraz nadzór nad zmianami modelu i jego wykorzystaniem[18].

Zidentyfikowane mechanizmy ryzyka związane z bezpieczeństwem danych i infrastruktury AI uwzględniono w zestawieniu przedstawionym w tabeli.

Ryzyka kompetencyjne i organizacyjne

Nadmierne poleganie na systemie

System, który przez dłuższy czas działa poprawnie, przestaje być podważany, a wynik modelu staje się domyślnym punktem odniesienia. Mechanizm nadmiernego polegania na systemach wspomagania decyzji jest szeroko opisywany w literaturze[4, 5] jako *automation bias* oraz *over-reliance* (nadmierne poleganie). Prowadzi on do przeniesienia części oceny sytuacji operacyjnej na algorytm.

Table. Key risk mechanisms associated with the use of artificial intelligence systems in a chemical plant

Tabela. Kluczowe mechanizmy ryzyka związane z wykorzystaniem systemów sztucznej inteligencji w instalacji chemicznej (*Źródło: opracowanie syntetyczne na podstawie literatury cytowanej w artykule*)

Obszar	Mechanizm	Skutek dla prowadzenia procesu	Wymagany nadzór
Zakres stosowalności modelu	model stosowany poza zakresem danych treningowych	rekomendacja nieadekwatna do aktualnego reżimu pracy; przesunięcie procesu w obszar słabo rozpoznany	walidacja w stanach przejściowych; jednoznaczne określenie zakresu stosowania
Utrata trafności w czasie	*data drift/concept drift* (zmiana rozkładu danych oraz zależności między zmiennymi procesowymi)	stopniowa degradacja jakości decyzji bez wyraźnego sygnału błędu	monitoring jakości predykcji; okresowa rewalidacja modelu
Integracja z systemem sterowania	wpływ rekomendacji modelu na parametry krytyczne procesu	zmniejszenie marginesu operacyjnego; niezamierzona zmiana poziomu ryzyka	traktowanie wdrożenia jako zmiany technologicznej; formalna ocena wpływu na bezpieczeństwo
Nadmierne poleganie na systemie (*automation bias*)	ograniczenie niezależnej weryfikacji decyzji	osłabienie czujności operatorskiej; opóźniona reakcja w sytuacji granicznej	utrzymanie realnego nadzoru człowieka; szkolenia
Odpowiedzialność	brak jednoznacznego właściciela modelu i jego aktualizacji	rozmycie odpowiedzialności w przypadku zdarzenia	wyznaczenie technicznego właściciela systemu AI; jasny podział ról
Bezpieczeństwo danych i infrastruktury AI	manipulacja danymi treningowymi lub zakłócenie danych procesowych	rekomendacje nieadekwatne do rzeczywistego stanu instalacji	kontrola integralności danych oraz weryfikacja źródeł informacji

W praktyce obserwowana jest również różnica pokoleniowa w sposobie korzystania z systemów AI. Starsi operatorzy zdobywali doświadczenie bez wsparcia narzędzi predykcyjnych, ucząc się reakcji procesu na podstawie obserwacji i konsekwencji wcześniejszych decyzji. Młodsze pokolenie wchodzi do pracy w środowisku, w którym system AI stanowi standard, a rekomendacja modelu jest dostępna od początku. W efekcie sposób budowania doświadczenia technologicznego ulega zmianie. Operatorzy bardziej doświadczeni częściej kwestionują wynik, natomiast mniej doświadczeni mogą traktować go jako punkt wyjścia do dalszej oceny. Jeżeli organizacja nie wymaga niezależnego uzasadnienia decyzji i nie wspiera transferu wiedzy, zależność od modelu rośnie, a rozwój kompetencji technologicznych może nie nadążać za zmianą sposobu pracy. W sytuacjach granicznych różnica ta staje się szczególnie istotna. System może nie rozpoznać odchylenia, które dla doświadczonego operatora stanowiłoby sygnał ostrzegawczy. Jeżeli w zespole dominuje nawyk polegania na modelu, reakcja może być opóźniona. W jednym z opisanych wdrożeń predykcyjnego utrzymania ruchu w przemyśle ciężkim podstawowy algorytm detekcji anomalii generował znaczną liczbę fałszywych alarmów. Skutkowało to częstą koniecznością weryfikacji wskazań przez personel oraz ryzykiem stopniowego spadku zaufania do systemu. Dopiero po modyfikacji algorytmu i wprowadzeniu dodatkowych mechanizmów filtrujących ograniczono liczbę fałszywych alarmów (*false positives*) średnio o ok. 90% względem pierwotnego rozwiązania[17]. Przykład ten pokazuje, że nawet technicznie poprawny model może w praktyce generować dodatkowe obciążenie informacyjne. Nadmiar sygnałów ostrzegawczych nie zwiększa bezpieczeństwa, lecz zmienia sposób reagowania zespołu i wpływa na czujność operacyjną. W literaturze dotyczącej zarządzania ryzykiem AI podkreśla się, że utrzymanie realnego nadzoru człowieka nad systemem wymaga nie tylko formalnej możliwości ingerencji, lecz także kompetencji pozwalających kwestionować wynik algorytmu[4, 6].

Odpowiedzialność

Wprowadzenie systemu AI nie zmienia formalnej odpowiedzialności za prowadzenie instalacji. Za decyzje operacyjne nadal odpowiada zespół prowadzący proces[6]. W praktyce model bywa jednak projektowany i wdrażany poza zespołem operacyjnym, np. przez dział IT, analityków danych lub zewnętrznego dostawcę. Zespół operacyjny korzysta z narzędzia, lecz nie zawsze zna założenia, na których zostało zbudowane ani zakres, w jakim było weryfikowane. W przypadku błędnej decyzji przyczyna bywa trudna do jednoznacznego wskazania: czy zawiodła ocena człowieka, czy model wygenerował nietrafną rekomendację, czy system został wprowadzony bez pełnej weryfikacji? W praktyce elementy te są ze sobą powiązane. Formalna odpowiedzialność pozostaje po stronie zespołu operacyjnego, natomiast faktyczny wpływ na przebieg procesu decyzyjnego może być rozproszony. Jeżeli nie jest jasno określone, kto nadzoruje model, kto odpowiada za jego aktualizację oraz kto może wstrzymać jego użycie, odpowiedzialność zaczyna się rozmywać. W dokumentach dotyczących zarządzania ryzykiem systemów AI podkreśla się konieczność jednoznacznego przypisania ról, odpowiedzialności i mechanizmów nadzoru w relacji człowiek-system[5]. W praktyce oznacza to potrzebę wyznaczenia właściciela modelu (*model owner*), odpowiedzialnego za nadzór, aktualizację, okresową weryfikację oraz możliwość wstrzymania jego użycia w przypadku wątpliwości co do adekwatności predykcji[4]. W środowisku instalacji chemicznej odpowiedzialność za kluczowe decyzje operacyjne jest przypisywana do konkretnych ról i osób dyżurnych.

Dlatego model, który wpływa na sposób prowadzenia procesu, powinien mieć jednoznacznie przypisanego właściciela oraz określone zasady nadzoru i eskalacji[4, 6]. W przeciwnym razie staje się elementem oddziałującym na ryzyko, za który nikt nie ponosi jasno określonej odpowiedzialności.

Rola człowieka w decyzjach operacyjnych

Decyzja operacyjna w instalacji przemysłowej rzadko wynika z pojedynczego sygnału. Parametry procesu mogą mieścić się w dopuszczalnym zakresie, mimo że pojawiają się symptomy wskazujące na możliwe odchylenia. Zmiany bywają subtelne: inna dynamika odpowiedzi układu, wolniejsza stabilizacja po korekcie nastawy lub nietypowy przebieg rozruchu. W takich sytuacjach interpretacja danych wymaga nie tylko analizy wartości liczbowych, lecz także zrozumienia kontekstu pracy instalacji.

Modele analityczne identyfikują zależności statystyczne między zmiennymi procesowymi na podstawie danych zapisanych w systemach informatycznych[19]. Ocena sytuacji operacyjnej uwzględnia natomiast również kontekst pracy procesu, doświadczenie operatora oraz wiedzę o ograniczeniach aparatury. Operator pamięta wcześniejsze zakłócenia, zna ograniczenia aparatury pomiarowej oraz potrafi rozpoznać zachowania procesu, które nie znajdują odzwierciedlenia w danych historycznych. W praktyce wiele decyzji podejmowanych jest przy niepełnej informacji, a ocena sytuacji opiera się na łączeniu rozproszonych sygnałów oraz doświadczeniu zdobytym w trakcie eksploatacji instalacji.

Podejście to jest szczególnie istotne w sytuacjach nietypowych, gdy proces odbiega od standardowych warunków pracy. Dane historyczne nie zawsze obejmują wszystkie możliwe scenariusze operacyjne, a część zjawisk ma charakter incydentalny lub trudny do jednoznacznego pomiaru. W takich przypadkach rola człowieka polega na ocenie wiarygodności dostępnych informacji oraz na rozważeniu możliwych konsekwencji podejmowanych działań.

Z tego względu systemy oparte na sztucznej inteligencji powinny być traktowane przede wszystkim jako narzędzia wspomagające analizę danych i identyfikację potencjalnych zależności. Mogą one porządkować duże zbiory informacji i wskazywać sygnały, które trudno byłoby zauważyć w tradycyjnej analizie, jednak nie zastępują oceny sytuacji procesowej dokonywanej przez człowieka. Zachowanie realnej kontroli operatora nad procesem oraz nad interpretacją rekomendacji systemu jest jednym z podstawowych warunków odpowiedzialnego wykorzystania narzędzi AI w środowisku przemysłowym, co podkreślają również dokumenty dotyczące zarządzania ryzykiem technologii cyfrowych[3, 4].

Tak rozumiana współpraca człowieka i systemów analitycznych oznacza, że algorytmy mogą rozszerzać zakres dostępnej informacji, natomiast ostateczna interpretacja sygnałów oraz decyzja operacyjna pozostają elementem praktyki eksploatacyjnej i odpowiedzialności personelu prowadzącego instalację.

W konsekwencji ważne staje się przełożenie powyższych założeń na praktykę wdrożeniową, obejmującą zarówno wymagania organizacyjne, jak i techniczne elementy utrzymania modeli w cyklu życia.

Perspektywa praktyki przemysłowej

W praktyce przemysłowej największe trudności związane z wdrażaniem systemów AI rzadko wynikają wyłącznie z konstrukcji modelu, co wskazują również przeglądy wdrożeń przemysłowych[9, 16]. Znacznie częściej dotyczą integracji z istniejącą infrastrukturą sterowania, jakości i spójności danych oraz dostępnych kompetencji organizacyjnych. W obiektach o długiej historii eksploatacji infrastruktura cyfrowa bywa niejednorodna, a dane historyczne nie zawsze odzwierciedlają pełny kontekst operacyjny, co ogranicza skuteczność narzędzi analitycznych[9, 16]. Dodatkowym wyzwaniem bywa rozbieżność celów między zespołami odpowiedzialnymi za rozwój narzędzi cyfrowych a zespołami operacyjnymi. Dla pierwszych kluczowa jest jakość predykcji modelu, dla drugich utrzymanie stabilności i bezpieczeństwa procesu. Presja na szybkie uzyskanie efektów wdrożeniowych może prowadzić do uproszczeń w zakresie walidacji, dokumentowania założeń oraz oceny wpływu na ryzyko. Z tego względu wdrażanie systemów AI w środowisku wysokiego ryzyka wymaga nie tylko kompetencji analitycznych, lecz także jasno zdefiniowanych zasad, przypisania odpowiedzialności oraz objęcia zmian formalnymi procedurami zarządzania zmianą (MOC) i nadzorem nad rozwiązaniami AI[4, 8]. Przedstawione w artykule mechanizmy mają charakter zarówno technologiczny, jak i organizacyjny. W praktyce prowadzenia instalacji nie funkcjonują one oddzielnie, lecz nakładają się na siebie, wpływając na interpretację danych oraz sposób podejmowania decyzji operacyjnych. Ich oddziaływanie nie musi prowadzić do natychmiastowego zakłócenia procesu; częściej skutkuje stopniowym zmniejszaniem marginesu operacyjnego i zmianą sposobu oceny sytuacji technologicznej. W celu uporządkowania zidentyfikowanych obszarów przedstawiono ich syntetyczne zestawienie w tabeli, odnosząc je do prowadzenia procesu oraz wymaganego nadzoru technicznego i organizacyjnego. Zestawienie to stanowi podsumowanie kluczowych punktów kontroli przy wdrażaniu systemów sztucznej inteligencji w instalacji chemicznej.

Podsumowanie i wnioski

Systemy sztucznej inteligencji stosowane w instalacjach chemicznych wpływają na sposób interpretacji danych oraz podejmowanie decyzji operacyjnych. Ich wdrożenie nie zmienia formalnej odpowiedzialności za prowadzenie instalacji, może jednak modyfikować praktyczny przebieg procesu decyzyjnego. Ryzyko związane z wykorzystaniem systemów AI nie ogranicza się wyłącznie do poprawności algorytmu. Obejmuje m.in. zakres stosowalności modeli, utrzymanie trafności w czasie, jakość danych wejściowych

oraz sposób integracji z istniejącą infrastrukturą techniczną. Istotne są również kompetencje zespołu oraz organizacja nadzoru nad systemem AI w całym jego cyklu życia. Model uczony na danych historycznych opisuje relacje obecne w danych przeszłych, podczas gdy proces przemysłowy funkcjonuje w warunkach zmiennych i nie w pełni przewidywalnych. W sytuacjach granicznych o bezpieczeństwie i jakości decyzji decydują nie tylko zależności zarejestrowane w bazie danych, lecz także doświadczenie zespołu oraz ocena kontekstu technologicznego. W środowisku wysokiego ryzyka systemy AI powinny być traktowane jako element systemu technicznego podlegający zasadom oceny, nadzoru oraz zarządzania zmianą, właściwym dla rozwiązań wpływających na sposób prowadzenia procesu. Odpowiedzialność za decyzje operacyjne i ich konsekwencje pozostaje po stronie człowieka. W praktyce oznacza to konieczność traktowania systemów AI jako elementów infrastruktury technicznej, podlegających formalnemu nadzorowi, ocenie ryzyka oraz zarządzaniu zmianą w całym cyklu życia[4, 6].

Otrzymano: 05-03-2026 *Zrecenzowano: 16-03-2026*
Zaakceptowano: 31-03-2026 *Opublikowano: 25-05-2026*

LITERATURA

[1] European Commission, Ethics Guidelines for Trustworthy AI, European Commission, Bruksela 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, dostęp 25.03.2026 r.
[2] NIST, AI Risk Management Framework (AI RMF 1.0), 2023, https://www.nist.gov/itl/ai-risk-management-framework, dostęp 25.03.2026 r.
[3] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji, *Dz.U. UE* L z 12.07.2024 r.
[4] S. Alon-Barkat, M. Busuioc, *J. Public Adm. Res. Theory* 2023, **33**, 153, DOI: 10.1093/jopart/muac007.
[5] A. Carter, *Process Saf. Environ. Prot*. 2023, **170**, 645, DOI: 10.1016/j.psep.2022.12.018.
[6] Center for Chemical Process Safety (CCPS), *Process safety metrics. Guide for leading and lagging indicators*, John Wiley & Sons, Hoboken 2021.
[7] IEC 61511-1:2016, *Functional safety. Safety instrumented systems for the process industry sector*.
[8] Center for Chemical Process Safety (CCPS), *Guidelines for management of change for process safety*, AIChE/Wiley, Hoboken 2008.
[9] Center for Chemical Process Safety (CCPS), *Guidelines for safe automation of chemical processes*, John Wiley & Sons / AIChE, Hoboken 2017.
[10] P. Daoutidis, J.H. Lee, S. Rangarajan i in., *Comput. Chem. Eng*. 2024, **181**, 108523, DOI: 10.1016/j.compchemeng.2023.108523.
[11] M. Mowbray, M.J.M. Simmons, D.A. Keen, *React. Chem. Eng*. 2022, **7**, 1515, DOI: 10.1039/D1RE00541C.
[12] U.S. Chemical Safety and Hazard Investigation Board (CSB), Investigation Report: BP Texas City Refinery Explosion and Fire, Report No. 2005-04-I-TX, 2007.
[13] D. Hendrycks, K. Gimpel, *Proc. Int. Conf. Learn. Represent*. (ICLR), 2017, https://arxiv.org/abs/1610.02136, dostęp 25.03.2026 r.
[14] A. Kahraman i in., *IEEE Access* 2022, **10**, 118143, DOI: 10.1109/ACCESS.2022.3210525.
[15] World Economic Forum, How digital transformation supports value creation in brownfield manufacturing, 2024, https://www.weforum.org/stories/2024/08/how-digital-transformation-supports-value-creation-in-brownfield-manufacturing/, dostęp 24.03.2026 r.
[16] S. Sekiou i in., *Comput. Chem. Eng*. 2026, **204**, 109378, DOI: 10.1016/j.compchemeng.2025.109378.
[17] M. Hermansa, M. Kozielski, M. Michalak i in., *Sensors* 2022, **22**, 226, DOI: 10.3390/s22010226.
[18] A. Vassilev, A. Oprea, A. Fordyce, H. Anderson, *Adversarial machine learning. A taxonomy and terminology of attacks and mitigations*, NIST, Gaithersburg 2024.
[19] S.J. Qin, *AIChE J*. 2014, **60**, 3092, DOI: 10.1002/aic.14523.