Diagnostyka operatora w przemyśle procesowym i nowa rola AI

dr inż. Paweł Wnuk, prof. dr hab. inż. Jan Maciej Kościelny

Instytut Automatyki i Robotyki, Politechnika Warszawska

Najkrótsza teza tego tekstu brzmi: w nowoczesnym zakładzie przemysłowym operator powinien być diagnozowany z taką samą powagą, z jaką diagnozuje się zawór, czujnik, pętlę regulacji i sieć sterowania. Nie po to, aby odebrać człowiekowi sprawczość, lecz po to, aby szybciej wykrywać błędne działania, lepiej rozumieć ich przyczyny i ograniczać ryzyko przejścia instalacji w stan awaryjny.

Streszczenie

Błędy ludzkie pozostają jedną z najczęstszych przyczyn incydentów i poważnych awarii w przemyśle procesowym. Jednocześnie klasyczna diagnostyka techniczna przez lata koncentrowała się głównie na uszkodzeniach aparatury, urządzeń automatyki, pętli regulacji oraz - coraz częściej - na cyberatakach. Brakującą warstwą jest automatyczna detekcja i diagnostyka błędów operatora, czyli OEDD (Operator Errors Detection and Diagnosis). Niniejszy artykuł przedstawia tę koncepcję w języku inżynierskim, bez nadmiernego formalizmu: jako praktyczny element architektury bezpieczeństwa dużej instalacji procesowej.

Punktem wyjścia jest regułowe podejście do modelowania poprawnych akcji operatora. Zakłada ono, że nie trzeba budować globalnego modelu człowieka-operatora. Wystarczy wybrać krytyczne sekwencje obsługi, opisać przesłanki ich uruchomienia, dopuszczalny czas reakcji i oczekiwane potwierdzenia skutków na obiekcie. W drugiej części artykułu pokazano, jak tę podstawę można rozszerzyć o AI: modele wykrywania anomalii, predykcję skutków działań, monitorowanie przeciążenia poznawczego oraz duże modele językowe (LLM), które mogą działać jako warstwa interpretacyjna nad procedurami, alarmami, logami operatorskimi i dokumentacją utrzymaniową. Ostatnia część omawia ryzyka stosowania AI: halucynacje, fałszywą pewność, utratę kalibracji zaufania, problemy po fine-tuningu, cyberbezpieczeństwo i odpowiedzialność organizacyjną.

Słowa kluczowe: diagnostyka operatora, OEDD, przemysł procesowy, systemy sterowania, bezpieczeństwo procesowe, AI, LLM, human-in-the-loop, diagnostyka techniczna.

1. Dlaczego diagnostyka operatora staje się tematem zarządczym

Wiele dużych zakładów przemysłowych ma dziś rozwiniętą diagnostykę aparatury, monitoring pętli regulacji, systemy alarmowe, zabezpieczenia SIS i procedury cyberbezpieczeństwa. To ogromny postęp w porównaniu z zakładami sprzed dwóch lub trzech dekad. A jednak w analizach poważnych awarii nadal regularnie wraca ten sam motyw: człowiek wykonał złą czynność, wykonał ją za późno, zignorował istotny sygnał, nie zobaczył kontekstu albo otrzymał tak dużo alarmów, że nie był w stanie właściwie ustalić priorytetów. Literatura dotycząca niezawodności człowieka wskazuje, że udział błędów ludzkich w wypadkach przemysłowych i operacjach krytycznych jest bardzo wysoki (Zarei i in., 2021; Abu Hawwach, 2021).

Nie chodzi przy tym o prostą tezę, że „operator jest winny”. W przemyśle procesowym błąd człowieka bardzo często jest skutkiem splotu czynników: złej jakości alarmów, niejednoznacznej procedury, presji czasu, zmęczenia, niepełnej informacji o stanie obiektu, braku potwierdzenia działania urządzenia wykonawczego albo konfliktu między lokalnym celem produkcyjnym a bezpieczeństwem procesu. Dlatego diagnostyka operatora nie może być traktowana jako system nadzoru personalnego w sensie administracyjnym. Powinna być projektowana jako warstwa bezpieczeństwa procesowego, która wykrywa niezgodność między wymaganym działaniem a działaniem rzeczywistym.

Dotychczasowa praktyka w wielu zakładach wygląda następująco: po awarii analizuje się archiwa procesu, listę alarmów, raport zmianowy i logi systemowe. Na tej podstawie rekonstruuje się, co zrobił operator i czy była to reakcja poprawna. OEDD odwraca tę logikę. Zamiast pytać po fakcie, co poszło źle, próbuje odpowiedzieć w czasie rzeczywistym: czy aktualna akcja operatora jest dopuszczalna, czy została wykonana w odpowiednim momencie, czy poprzedziły ją właściwe przesłanki i czy obiekt potwierdził jej skutek.

Dla dyrektorów technicznych i szefów automatyki kluczowy wniosek jest prosty: diagnostyka operatora nie jest kolejnym modułem raportowym. Jest brakującym ogniwem między klasyczną automatyką, bezpieczeństwem procesowym, cyberbezpieczeństwem i zarządzaniem kompetencjami operacyjnymi. Jeżeli zakład mierzy stan pompy, zaworu i czujnika, ale nie potrafi automatycznie ocenić krytycznej sekwencji działań człowieka, to pozostawia niezamkniętą jedną z najważniejszych ścieżek ryzyka.

Demografia tylko pogłębia presję na monitoring kompetencji. Według raportu Deloitte i The Manufacturing Institute „Creating pathways for tomorrow's workforce today: Beyond reskilling in manufacturing" (4 maja 2021), luka kompetencyjna w przemyśle USA może doprowadzić do 2,1 mln nieobsadzonych stanowisk do 2030 r., kosztem rzędu 1 bln USD rocznie. Polska petrochemia i energetyka stoją przed analogicznym problemem retentu doświadczonych operatorów.

2. OEDD: czym różni się od klasycznej diagnostyki

Klasyczna diagnostyka uszkodzeń, określana jako FDD lub FDI, rozwija się od wielu dekad. Jej podstawowa idea polega na porównaniu bieżącego zachowania procesu z zachowaniem oczekiwanym. Różnica między sygnałem mierzonym a sygnałem wynikającym z modelu tworzy residuum, które po ocenie może stać się sygnałem diagnostycznym. Jeżeli zestaw sygnałów diagnostycznych odpowiada określonemu wzorcowi, system wskazuje prawdopodobne uszkodzenie (Blanke i in., 2004; Korbicz i in., 2004; Isermann, 2006).

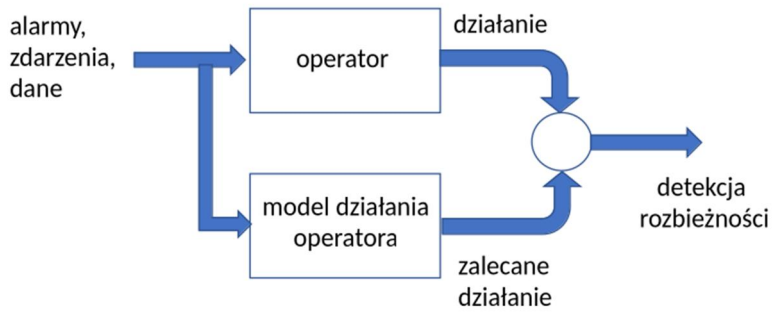
Analogiczna idea może zostać zastosowana do działań operatora. Zamiast porównywać jedynie obiekt z modelem fizycznym, porównujemy rzeczywistą akcję operatora z modelem poprawnej akcji. Modelem nie musi być skomplikowany opis psychologiczny człowieka. Wystarczy model cząstkowy, który mówi: w określonych warunkach procesowych operator powinien wykonać daną akcję, w określonym czasie i z określonym skutkiem. Jeżeli akcja jest wykonana bez przesłanek, zbyt późno, w złej kolejności albo nie daje potwierdzenia na obiekcie, pojawia się symptom błędu operatora lub uszkodzenia urządzenia realizującego tę akcję.

Tu kryje się praktyczna siła OEDD. System nie próbuje oceniać człowieka jako całości. Nie stwierdza, czy operator „dobrze pracuje” albo czy „ma odpowiednie kompetencje”. Ocenia

konkretne, zdefiniowane wcześniej działanie w konkretnym kontekście technologicznym. Dzięki temu można zacząć od najważniejszych sekwencji: rozruchów, odstawień, przełączeń trybów pracy, ręcznego sterowania elementami wykonawczymi, reakcji na alarmy krytyczne, obejść zabezpieczeń oraz działań wykonywanych rzadko, ale o wysokim potencjale skutków.

W praktyce OEDD wymaga trzech klas sygnałów. Po pierwsze, sygnałów opisujących stan procesu: pomiarów, alarmów, zdarzeń, stanów logicznych i sygnałów sterujących. Po drugie, logów dokumentujących ingerencje operatora: zmiany wartości zadanych, przełączenia trybów, komendy otwarcia lub zamknięcia, potwierdzenia alarmów, ręczne obejścia i decyzje proceduralne. Po trzecie, sygnałów potwierdzających skutek działania: stanu urządzenia wykonawczego, czujników krańcowych, zmian przepływu, ciśnienia, temperatury albo innych zmiennych, które dokumentują, że akcja rzeczywiście została zrealizowana.

Jeżeli którejś z tych klas brakuje, jakość diagnozy spada. Zakład może wykryć anomalię, ale nie potrafi rozstrzygnąć, czy przyczyną był operator, uszkodzony zawór, błędny sygnał zwrotny, opóźnienie komunikacji czy cyberatak. Dlatego diagnostyka operatora zaczyna się nie od zakupu algorytmu, lecz od inwentaryzacji danych: co jest rejestrowane, z jaką rozdzielczością czasową, czy logi operatorskie są kompletne, czy mają spójne znaczniki czasu i czy można je powiązać z danymi procesowymi.



Rys. 1. Schemat detekcji błędnych działań operatorów: rzeczywiste działanie operatora jest porównywane z modelem działania oczekiwanego w danym stanie procesu.

3. Reguły zamiast czarnej skrzynki

W świecie AI modne są dziś modele uczone na danych. W diagnostyce operatora nie należy jednak od nich zaczynać. Powód jest bardzo prozaiczny: w wielu instalacjach nie istnieją duże, dobrze opisane zbiory danych obejmujące poprawne i błędne działania operatorskie we wszystkich stanach awaryjnych. Co więcej, im bezpieczniejsza instalacja, tym mniej ma danych o realnych awariach. Paradoksalnie więc najlepsze zakłady mogą mieć najgorszy materiał do uczenia modeli czysto statystycznych.

Dlatego pierwszą warstwą OEDD powinny być modele regułowe, tworzone z wiedzy technologów, automatyków, doświadczonych operatorów, dokumentacji producentów i zatwierdzonych instrukcji. Reguła może mieć formę prostą: jeżeli spełniony jest warunek technologiczny A, aktywne jest zezwolenie B, a alarm C ma określony stan, to w czasie T operator

powinien wykonać akcję D. Następnie w czasie T2 powinien pojawić się skutek E. Jeżeli akcja D wystąpi bez A i B, jest nieuzasadniona. Jeżeli A i B występują, a D nie pojawia się w czasie T, mamy brak wymaganej akcji. Jeżeli D występuje, ale nie pojawia się E, należy rozważyć uszkodzenie urządzenia wykonawczego albo błąd w torze potwierdzenia.

Takie podejście ma kilka zalet menedżerskich. Po pierwsze, jest audytowalne: każdy może zobaczyć, dlaczego system zgłosił błąd. Po drugie, jest wdrażalne od początku życia instalacji, bez czekania na historię awarii. Po trzecie, pozwala łatwo uzgadniać odpowiedzialność między działami: technologia definiuje warunki procesu, automatyka wskazuje dostępne sygnały, utrzymanie ruchu opisuje potwierdzenia urządzeń, a operatorzy weryfikują realność sekwencji.

Po czwarte, reguły są odporne na problem, który często utrudnia wdrożenia AI w przemyśle: przenoszenie modelu między instalacjami. Model neuronowy wytrenowany na jednej instalacji może działać słabo w innej, bo zmieniają się skale, dynamika, konfiguracja alarmów i praktyka obsługi. Reguła proceduralna również wymaga dostosowania, ale jej sens jest zrozumiały. Można ją przenieść i sprawdzić inżyniersko.

Nie oznacza to, że reguły wystarczą zawsze. Ich słabością jest konieczność ręcznego utrzymywania, ryzyko niepełnego pokrycia sytuacji nietypowych oraz trudność opisu działań silnie zależnych od kontekstu. Z tego powodu rozsądna architektura nie przeciwstawia reguł i AI. Najlepiej traktować reguły jako twardą, wyjaśnialną warstwę podstawową, a AI jako warstwę rozszerzającą: wykrywającą nietypowe wzorce, przewidującą konsekwencje, streszczającą sytuację i wspierającą operatora w interpretacji.

Klasa informacji	Przykłady sygnałów	Znaczenie dla diagnozy
Stan procesu	pomiar ciśnienia, temperatury i przepływu; alarmy; zezwolenia technologiczne; stany logiczne	określa, czy akcja operatora była wymagana lub dopuszczalna
Akcja operatora	polecenie otwarcia/zamknięcia, zmiana wartości zadanej, przełączenie trybu, potwierdzenie alarmu	pozwala stwierdzić, co operator faktycznie zrobił i kiedy
Skutek akcji	krańcówka zaworu, potwierdzenie pracy pompy, zmiana przepływu, ciśnienia lub temperatury	pozwala odróżniać błąd operatora od uszkodzenia urządzenia albo toru pomiarowego

Tabela 1. Minimalny zestaw informacji potrzebnych do sensownej diagnostyki działań operatora.

4. Co właściwie można diagnozować

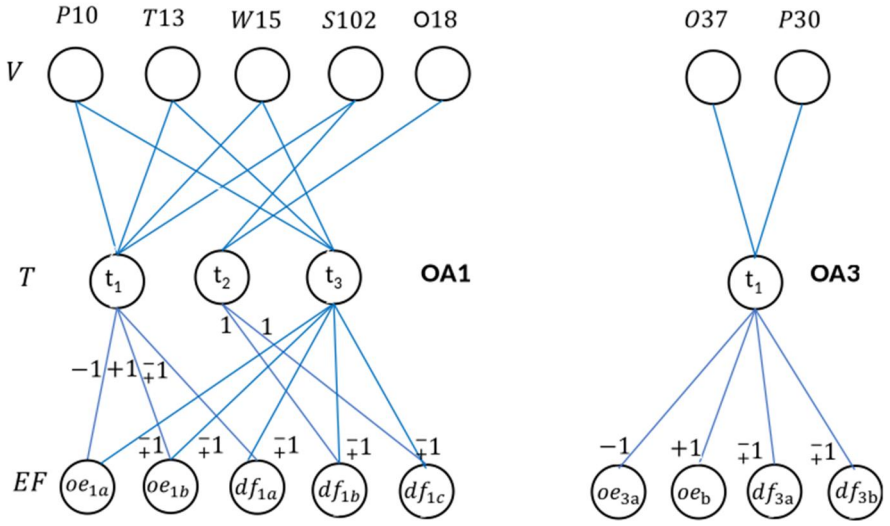
OEDD staje się jasne dopiero wtedy, gdy pokaże się typowe klasy błędów. Pierwsza klasa to nieuzasadnione działanie operatora: na przykład otwarcie zaworu, uruchomienie pompy albo zmiana struktury regulacji bez wymaganych przesłanek technologicznych. Druga klasa to brak działania: operator powinien wykonać akcję po spełnieniu warunków, ale nie robi tego w wymaganym czasie. Trzecia klasa to działanie spóźnione albo przedwczesne. Czwarta klasa to działanie o złym parametrze: zbyt duża zmiana wartości zadanej, przejście na tryb ręczny bez ograniczeń albo sterowanie elementem wykonawczym w kierunku pogarszającym odchyłkę.

Osobną grupą są działania naruszające strukturę sterowania. Wprowadzenie dodatniego sprzężenia zwrotnego przez nieprawidłową zmianę trybu pracy, rozpięcie kaskady bez uzasadnienia, obejście blokady albo utrzymanie elementu wykonawczego w pozycji krańcowej

może prowadzić do szybkiego pogorszenia stabilności procesu. W takich przypadkach czas detekcji ma krytyczne znaczenie: system nie powinien czekać, aż skutki błędu rozwiną się w danych procesowych. Powinien wykryć błąd już na poziomie samej akcji i jej kontekstu.

Ważne jest także rozróżnienie błędu operatora i uszkodzenia urządzenia. Jeżeli operator wydał poprawne polecenie zamknięcia zaworu, ale krańcówka nie potwierdza zamknięcia, nie można automatycznie mówić o błędzie człowieka. Możliwe jest uszkodzenie zaworu, napędu, krańcówki, okablowania lub toru komunikacji. Jeżeli natomiast zawór został zamknięty bez warunków dopuszczenia, błąd leży po stronie akcji operatorskiej albo procedury, która do tej akcji dopuściła. Dobra diagnostyka operatora nie zastępuje diagnostyki technicznej; ona musi z nią współpracować.

To rozróżnienie ma bezpośredni wymiar organizacyjny. Alarm „błąd operatora" jest bardzo wrażliwy. Źle zaprojektowany może budzić opór załogi i prowadzić do ukrywania problemów. Dlatego komunikat powinien być formułowany technicznie: „akcja niezgodna z warunkami procedury", „brak wymaganej akcji w czasie", „brak potwierdzenia skutku akcji", „akcja nierozróżnialna z uszkodzeniem urządzenia wykonawczego". Taki język przenosi rozmowę z personalnej oceny na analizę procesu i systemu pracy.



Rys. 2. Przykład grafowego ujęcia zależności między akcjami operatora, możliwymi błędami i uszkodzeniami urządzeń wykonawczych.

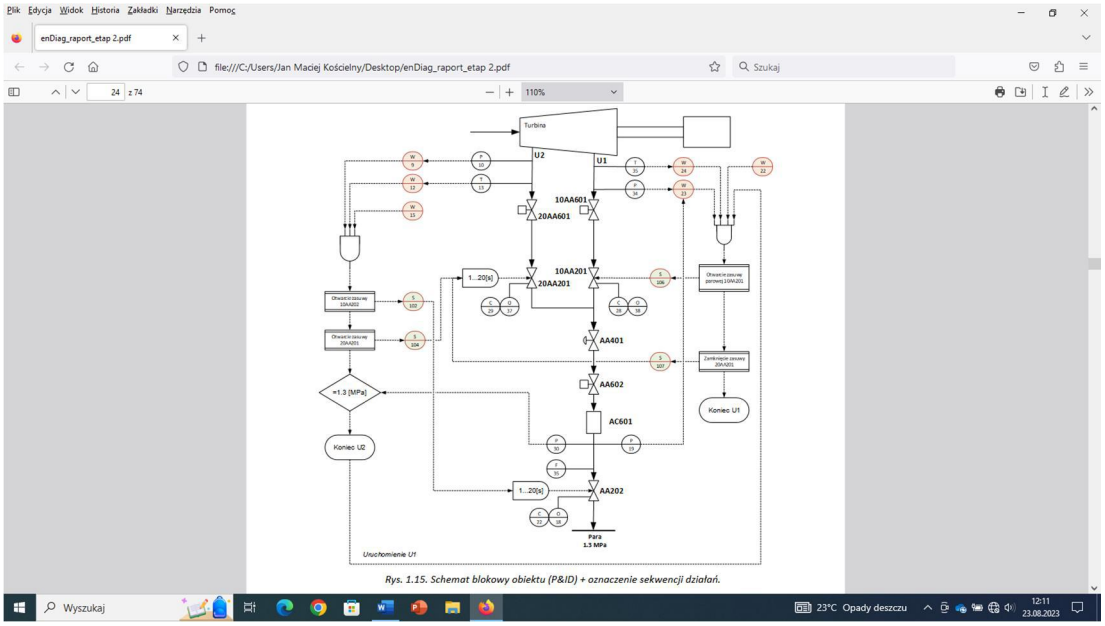
5. Studium przypadku: uruchamianie upustów turbiny

Jednym z dobrych przykładów pokazujących sens OEDD jest sekwencja uruchamiania upustów turbiny parowej. Z punktu widzenia operatora jest to procedura techniczna: trzeba sprawdzić warunki dopuszczenia, wykonać odpowiednie akcje w określonej kolejności i obserwować skutki. Z punktu widzenia systemu diagnostycznego jest to idealny przypadek do modelowania regułowego, ponieważ występują jasno określone warunki rozpoczęcia, polecenia sterujące, potwierdzenia i mierzalne skutki procesowe.

W uproszczeniu reguła dla takiej procedury może mówić: jeżeli spełnione są warunki pracy upustu, ciśnienie przekracza wymagany próg, a temperatura mieści się w dopuszczalnym zakresie, operator powinien wygenerować sygnał otwarcia. Następnie system powinien zaobserwować potwierdzenie stanu urządzenia i odpowiednią reakcję zmiennych procesowych. Jeżeli sygnał otwarcia pojawił się mimo niespełnienia warunków, mamy nieuzasadnioną akcję. Jeżeli warunki są spełnione, ale sygnał nie pojawia się w oknie czasowym, mamy brak akcji. Jeżeli sygnał pojawia się, lecz obiekt nie reaguje, pojawia się hipoteza uszkodzenia urządzenia albo toru potwierdzenia.

Ten przykład jest istotny nie dlatego, że dotyczy konkretnej turbiny. Jest istotny dlatego, że pokazuje wzorzec możliwy do przeniesienia na inne obiekty: uruchomienie pompy rezerwowej, przełączenie zasilania, odstawienie fragmentu instalacji, przejście pętli regulacji w tryb ręczny, ręczne sterowanie zaworem przy rozruchu, przygotowanie układu do remontu albo reakcję na alarm krytyczny. W każdym z tych przypadków można opisać warunki, akcję, czas i potwierdzenie skutku.

Z punktu widzenia zarządzania utrzymaniem ruchu dodatkową korzyścią jest dokumentowanie „prawie błędów". System OEDD może rejestrować sytuacje, w których akcja była spóźniona, ale nie doprowadziła do awarii; albo akcja była poprawna, lecz potwierdzenie urządzenia pojawiło się z opóźnieniem. Takie dane są bardzo wartościowe w szkoleniu operatorów, racjonalizacji alarmów, przeglądzie procedur i analizie niezawodności urządzeń wykonawczych. Zakład zaczyna widzieć nie tylko awarie, lecz także słabe sygnały poprzedzające awarie.



Rys. 3. Uproszczony schemat obiektu i sekwencji działań operatora dla przykładu uruchamiania upustów turbiny.

6. Nowa rola AI: od wykrywania akcji do rozumienia sytuacji

W ostatnich latach badania nad AI dla błędu ludzkiego zaczęły wychodzić poza klasyczne podejście „wykryj anomalię w danych". Przeglądy literatury wskazują cztery nurty: opisowy, predykcyjny, preskrypcyjny i generatywny (Gursel i in., 2025). W praktyce oznacza to, że AI może nie tylko wykrywać, że operator zrobił coś nietypowego, lecz także opisywać sytuację,

przewidywać skutki planowanej akcji, podpowiadać bezpieczniejsze działanie i generować czytelne wyjaśnienia dla człowieka.

Pierwszy obszar to uczenie maszynowe do wykrywania anomalii w działaniach i danych operatorskich. Prace dotyczące energetyki jądrowej pokazały, że modele nienadzorowane, w tym sieci typu GAN, mogą wykrywać niespójności między danymi z sensorów a danymi ręcznie wprowadzanymi przez personel (Gursel i in., 2023). Dla przemysłu procesowego ważna jest tu nie sama technika GAN, lecz szersza idea: można analizować spójność między tym, co mówi obiekt, tym, co rejestruje system sterowania, i tym, co deklaruje lub wprowadza człowiek.

Drugi obszar to predykcja skutków działań operatora. W instalacjach krytycznych operator często nie potrzebuje kolejnego alarmu, ale odpowiedzi na pytanie: co stanie się, jeśli teraz wykonam tę akcję? Modele predykcyjne mogą prognozować trendy parametrów po określonych działaniach operatorskich i wspierać ocenę planu działań. Taka funkcja jest szczególnie przydatna w stanach nienormalnych, gdy dynamika procesu jest szybka, a intuicja operatora bywa obciążona stresem i niepełną informacją.

Trzeci obszar to nadzór nad stanem poznawczym operatora. Coraz więcej badań wykorzystuje okulografię, EEG i inne sygnały fizjologiczne do oceny obciążenia poznawczego, uwagi i sposobu pozyskiwania informacji. Połączenie eye-trackingu i EEG pozwala klasyfikować poziom obciążenia operatorów nastawni, a analiza entropii spojrzenia może wskazywać przeciążenie w sytuacjach nienormalnych. Wdrożeniowo jest to trudniejszy obszar, bo dotyka prywatności i akceptacji społecznej, ale jego znaczenie będzie rosło tam, gdzie błędna percepcja sytuacji jest równie groźna jak błędna akcja.

Czwarty obszar to walidacja procedur i koordynacja działań. AI może sprawdzać, czy sekwencja akcji jest zgodna z procedurą, czy nie pominięto kroku, czy kolejność jest dopuszczalna i czy warunki procesu uzasadniają przejście do następnego etapu. W bardziej zaawansowanym wariancie AI może pomagać rozstrzygać konflikty: na przykład, gdy szybkie przywrócenie produkcji konkuruje z minimalizacją ryzyka termicznego, mechanicznego albo środowiskowego. W tym obszarze działają także rozwiązania regułowe opisane w pierwszej części artykułu.

7. Duże modele językowe: nie sterownik, lecz warstwa interpretacyjna

Duże modele językowe (LLM) nie powinny być traktowane jako kolejny algorytm wpięty bezpośrednio w sterowanie. Ich największa wartość leży gdzie indziej: potrafią łączyć informacje z wielu źródeł, które w zakładzie często żyją osobno. Historian przechowuje szeregi czasowe. DCS/SCADA zapisuje alarmy i komendy. CMMS ma historię prac utrzymaniowych. Operatorzy piszą raporty zmianowe. Producenci dostarczają instrukcje. Technolodzy mają procedury rozruchu i odstawienia. LLM może stać się interfejsem do tej rozproszonej wiedzy, o ile jest osadzony w lokalnych, zatwierdzonych źródłach.

Pierwsze zastosowanie to „strażnik procedury". Model językowy, działający w architekturze RAG, może porównywać aktualną sekwencję z obowiązującą instrukcją i wskazywać brakujący krok, zbyt wczesną akcję, niezgodność warunków dopuszczenia albo brak potwierdzenia skutku. Ważne jest, aby model nie generował procedury z pamięci, lecz cytował konkretny punkt dokumentu zakładowego i powiązane sygnały procesowe. Wtedy operator widzi nie tylko

rekomendację, ale również jej podstawę. W tym sensie zastosowanie AI ma dwojaki sens – z jednej strony jest to system doradczy, z drugiej – kontrolny. Jednakże, ze względu na specyfikę modeli LLM (czas działania, propabalistyczny charakter odpowiedzi) – w przeciwieństwie do wcześniej opisanych rozwiązań OEDD – modele AI tego typu raczej nie powinny działać w sposób zautomatyzowany, blokujący / zmieniający decyzje operatora.

Drugie zastosowanie to tłumaczenie złożonej sytuacji technicznej. W stanie alarmowym problemem nie jest brak danych, lecz ich nadmiar. LLM może streścić lawinę alarmów, wskazać pierwsze zdarzenia w ciągu przyczynowym, oddzielić alarmy pierwotne od wtórnych i zaproponować listę hipotez do sprawdzenia. Dla mistrza zmiany lub dyrektora technicznego może przygotować krótką informację: co się stało, jakie działania wykonano, co jest potwierdzone, czego nie wiemy i jakie ryzyka pozostają.

Trzecie zastosowanie to interfejs do diagnostyki technicznej. Model językowy może prowadzić rozmowę z automatykiem lub operatorem: „pokaż ostatnie nastawy dla tego zaworu", „porównaj przebieg z poprzednim rozruchem", „sprawdź, czy podobna sekwencja wystąpiła w ostatnich sześciu miesiącach", „jakie były prace UR na tym napędzie". W połączeniu z cyfrowym bliźniakiem albo bazą usterek może generować hipotezy przyczynowe i proponować kolejne kroki weryfikacji.

Czwarte zastosowanie to dokumentowanie incydentu i uczenie organizacji. Po zdarzeniu LLM może przygotować uporządkowany raport: chronologię sygnałów, akcje operatora, rozbieżności z procedurą, potwierdzenia skutków, decyzje podjęte przez zespół i otwarte pytania do przeglądu. Taki raport nie zastępuje analizy przyczyn źródłowych, ale skraca czas zebrania materiału i ułatwia rozmowę między produkcją, automatyką, UR i bezpieczeństwem procesowym.

Najważniejsza zasada brzmi jednak: LLM nie powinien być pierwszą linią bezpieczeństwa. Pierwszą linią pozostają klasyczne rozwiązania: SIS, twarde reguły, walidacja sygnałów i klasyczne algorytmy diagnostyczne. LLM jest drugą linią rozumienia sytuacji. Ma pomagać w interpretacji, komunikacji, wyszukiwaniu wiedzy i dokumentowaniu, ale nie powinien samodzielnie wykonywać działań wykonawczych w instalacji krytycznej.

Warstwa	Typowe narzędzia	Rola w OEDD
Deterministyczna	reguły proceduralne, interlocki, SIS, FDD, walidacja sensorów	szybka i audytowalna detekcja niezgodności oraz blokada działań niedopuszczalnych
Analityczna AI/ML	modele anomalii, predykcja trendów, klasyfikacja obciążenia poznawczego	rozszerzanie pola widzenia o wzorce trudne do zapisania w regułach
Językowa	LLM z RAG, lokalna baza wiedzy, raporty zmianowe, procedury	wyjaśnianie sytuacji, sprawdzanie procedur, dialog diagnostyczny i dokumentowanie incydentów
Organizacyjna	przeglądy procedur, szkolenia, analiza near-miss, zarządzanie zmianą	zamiana diagnoz w trwałą poprawę systemu pracy

Tabela 2. Przykładowa warstwowa architektura diagnostyki operatora z wykorzystaniem AI.

8. Ryzyka stosowania AI w nadzorze nad operatorem

Pierwsze ryzyko to fałszywa pewność. Modele AI, a szczególnie LLM, potrafią formułować odpowiedzi płynne, logiczne i przekonujące nawet wtedy, gdy są niepełne albo błędne. W biurze może to być kłopotliwa pomyłka. W nastawni może to być źródło decyzji prowadzącej do pogorszenia stanu instalacji. Dlatego każdy system generujący rekomendacje musi pokazywać poziom niepewności, źródła danych, ograniczenia oraz rozróżnienie między faktem, hipotezą i sugestią. W środowisku safety-critical brak tej transparentności jest wadą techniczną, a nie kosmetycznym problemem interfejsu.

Drugie ryzyko wiąże się z dostrajaniem modeli. Badania nad bezpieczeństwem LLM pokazują, że fine-tuning nawet na pozornie wąskich zadaniach może osłabiać wcześniejsze mechanizmy bezpieczeństwa lub prowadzić do nieoczekiwanych zachowań modelu (Qi i in., 2024; Betley i in., 2026). Dla przemysłu oznacza to, że model dostrojony do dokumentacji UR, alarmów i instrukcji nie może być traktowany jak zwykła aplikacja biurowa. Każda aktualizacja modelu, danych RAG lub promptów systemowych powinna podlegać zarządzaniu zmianą, testom regresyjnym i walidacji na scenariuszach awaryjnych.

Trzecie ryzyko to zła kalibracja zaufania. Jeżeli model często pomaga, operator może zacząć ufać mu nadmiernie. Jeżeli kilka razy się pomyli, załoga może go całkowicie odrzucić. Oba scenariusze są niebezpieczne. Badania nad human-AI teaming podkreślają znaczenie transparentności, wyjaśnialności i wspólnej świadomości sytuacyjnej (Endsley, 2023; Saghafian i in., 2025). System powinien więc uczyć właściwego korzystania z AI: kiedy model jest pomocny, kiedy wymaga potwierdzenia, a kiedy nie wolno na nim polegać.

Czwarte ryzyko jest organizacyjne i etyczne. Diagnostyka operatora może zostać odebrana jako narzędzie kontroli pracowników, a nie jako narzędzie bezpieczeństwa. Jeżeli wdrożenie będzie prowadzone bez udziału operatorów, bez jasnych zasad wykorzystania danych i bez rozdzielenia diagnostyki technicznej od oceny personalnej, może wywołać opór i pogorszyć kulturę bezpieczeństwa. Dlatego należy od początku komunikować, że celem OEDD jest wykrywanie niebezpiecznych sekwencji i słabych punktów systemu pracy, a nie automatyczne karanie ludzi.

Piąte ryzyko dotyczy cyberbezpieczeństwa. LLM podłączony do dokumentacji, historianów, systemów alarmowych i raportów zmianowych staje się atrakcyjnym celem. Trzeba chronić go przed prompt injection, manipulacją dokumentami źródłowymi, nieuprawnionym dostępem do danych i wyciekiem informacji o instalacji. Model nie powinien mieć bezpośrednich uprawnień wykonawczych, a jego odpowiedzi powinny być logowane, wersjonowane i możliwe do odtworzenia po incydencie.

9. Jak wdrażać OEDD i AI rozsądnie

Rozsądne wdrożenie warto zacząć od mapy ryzyka operatorskiego. Nie należy próbować objąć całego zakładu jednym projektem. Najpierw trzeba wybrać kilka sekwencji o wysokim znaczeniu: rozruch, odstawienie, przełączenie trybów, reakcja na alarm krytyczny albo operacja wykonywana rzadko, ale potencjalnie groźna. Dla każdej sekwencji należy opisać warunki dopuszczenia, wymagane akcje, dopuszczalne czasy, oczekiwane potwierdzenia i możliwe skutki błędu.

Drugi krok to audyt danych. Zakład powinien sprawdzić, czy rejestruje wszystkie akcje operatora, czy znaczniki czasu są spójne między systemami, czy alarmy są racjonalizowane, czy zachowane są dane o potwierdzeniach urządzeń wykonawczych i czy można połączyć historię procesu z logami operatorskimi. Bez tego OEDD szybko zamieni się w system generujący trudne do rozstrzygnięcia alarmy.

Trzeci krok to budowa reguł i ich przegląd z załogą. Najlepsze reguły powstają wtedy, gdy automatyk, technolog, operator i utrzymanie ruchu siedzą przy jednym stole. Technolog wie, kiedy akcja jest dopuszczalna. Automatyk wie, czy sygnały są dostępne. Operator wie, czy procedura jest realna w warunkach zmiany. Utrzymanie ruchu wie, jakie potwierdzenia bywają zawodne. Taka praca ma dodatkową wartość: porządkuje procedury i ujawnia luki, zanim powstanie system informatyczny.

Czwarty krok to wdrożenie pilotażowe bez automatycznego sankcjonowania. Przez pierwsze miesiące system powinien działać jako obserwator i generator raportów. Trzeba mierzyć liczbę wykrytych niezgodności, fałszywe alarmy, czas reakcji, wpływ na obciążenie operatorów i przydatność w analizach zmianowych. Dopiero po tej fazie można decydować, które komunikaty powinny trafić do alarmów bieżących, które do raportów, a które do szkoleń.

Piąty krok to dołączenie AI i LLM w sposób warstwowy. Modele ML mogą analizować anomalie i predykcje, a LLM może tworzyć warstwę dialogu i interpretacji. Nie należy jednak zaczynać od „asystenta, który odpowie na wszystko". Lepiej zdefiniować konkretne funkcje: streszczenie sekwencji alarmów, porównanie przebiegu z procedurą, wyszukanie podobnych incydentów, wygenerowanie listy kontrolnej lub przygotowanie raportu po zdarzeniu. Każda funkcja powinna mieć testy i jasne granice odpowiedzialności.

Szósty krok to utrzymanie systemu. Procedury się zmieniają, instalacje są modernizowane, alarmy są racjonalizowane, a praktyka operatorska ewoluuje. OEDD i AI muszą mieć właściciela procesowego, harmonogram przeglądów, repozytorium reguł, rejestr zmian i mechanizm wycofywania błędnych wersji. W przeciwnym razie system, który na początku poprawiał bezpieczeństwo, po kilku latach może stać się źródłem nieaktualnych komunikatów.

Powyższe rozważania powinny obejmować każdy rodzaj wdrożenia. Aktualnie wiodący dostawcy rozwiązań automatyki oferują rozwiązania klasy "AI assistant for operator", przykłady:

Honeywell — Experion Operations Assistant. Honeywell ogłosił 1 października 2024 r. w Houston wprowadzenie Experion Operations Assistant — explainable AI zintegrowanego z DCS Experion® PKS,

Emerson — Ovation AI-enabled Virtual Advisor. Emerson 15 lipca 2025 r. (Pittsburgh) wprowadził Ovation AI-enabled Virtual Advisor, opisany w oficjalnym komunikacie jako „first generative artificial intelligence (GenAI) advisor integrated into an automation system specifically designed for the power and water industries".

Yokogawa — OpreX + UptimeAI. 24 stycznia 2025 r. Yokogawa ogłosiła strategiczne partnerstwo z UptimeAI, integrując OpreX Asset Health Insights z modułami „AI Expert: Generative AI" oraz „AI Expert: Reliability & Process" — LLM-based agentami dla operatorów. Wcześniej Yokogawa wdrożyła autonomous control AI (FKDPP) komercyjnie w ENEOS Materials (2023).

Schneider Electric / AVEVA. Na targach Automate 2025 (Detroit, maj 2025) Schneider Electric ogłosił Industrial Copilot zintegrowany z EcoStruxure Automation Expert Platform, oparty o Microsoft Azure AI Foundry.

Podobne doniesienia pojawiają się wśród kolejnych dostawców – możemy się spodziewać że stoimy w przededniu kolejnej rewolucji.

10. Co to zmienia dla dyrektora i szefa automatyki

Dla dyrektora technicznego OEDD jest sposobem na przejście od deklaratywnego bezpieczeństwa do mierzalnej kontroli krytycznych sekwencji. Wiele zakładów ma procedury, ale nie wie, jak często są realizowane z odchyleniami, które odchylenia są niegroźne, a które powtarzają się jako słabe sygnały. Diagnostyka operatora pozwala przekształcić procedurę w mierzalny model działania.

Dla szefa automatyki OEDD oznacza nowe wymagania wobec systemu sterowania. Ważne stają się kompletne logi operatorskie, synchronizacja czasu, jakość sygnałów potwierdzających, możliwość wiązania alarmów z akcjami i łatwy dostęp do danych historycznych. To często mniej spektakularne niż wdrożenie AI, ale bez tego AI nie będzie miała stabilnego gruntu.

Dla utrzymania ruchu OEDD jest źródłem informacji o urządzeniach wykonawczych. Jeżeli poprawna komenda regularnie nie daje oczekiwanego potwierdzenia albo daje je z opóźnieniem, może to wskazywać na problem z napędem, zaworem, czujnikiem krańcowym lub komunikacją. W ten sposób diagnostyka operatora wspiera również diagnostykę techniczną.

Dla operatorów dobrze wdrożony system może być realnym wsparciem. Nie zastępuje doświadczenia, ale pomaga w sytuacjach przeciążenia: przypomina warunki, wskazuje brakujące potwierdzenia, porządkuje alarmy, pokazuje możliwe konsekwencje i dokumentuje przebieg zdarzenia. Warunkiem jest jednak zaufanie zbudowane przez transparentność i udział operatorów w projektowaniu reguł.

Największa zmiana kulturowa polega na tym, że błąd operatora przestaje być wyłącznie tematem dochodzenia po incydencie. Staje się normalnym przedmiotem diagnostyki technicznej: wykrywalnym, analizowalnym, porównywalnym i możliwym do ograniczania poprzez lepsze procedury, lepsze alarmy, lepsze interfejsy i lepsze szkolenie.

11. Wnioski

Przemysł procesowy od lat rozwija diagnostykę urządzeń, pętli regulacji i cyberbezpieczeństwa. Czas dodać do tej architektury diagnostykę operatora. Nie jako narzędzie personalnej kontroli, lecz jako brakującą warstwę bezpieczeństwa, która widzi relację między stanem procesu, akcją człowieka i skutkiem na obiekcie.

Najbardziej praktycznym fundamentem OEDD są dziś modele regułowe. Są zrozumiałe, audytowalne i możliwe do uruchomienia bez wieloletnich zbiorów danych awaryjnych. AI rozszerza ten fundament tam, gdzie reguły są zbyt sztywne: w wykrywaniu wzorców, predykcji skutków, ocenie przeciążenia poznawczego i syntezie rozproszonej wiedzy. LLM mogą wnieść szczególnie dużą wartość jako warstwa interpretacyjna, ale nie powinny zastępować twardych zabezpieczeń ani wykonywać działań w instalacji krytycznej.

Najdojrzalszy kierunek na najbliższe lata to więc nie „AI zamiast operatora", lecz AI wspierająca operatora i nadzorująca jakość krytycznych działań operatorskich. W takim modelu człowiek pozostaje odpowiedzialnym uczestnikiem procesu, automatyka dostarcza twardych danych i zabezpieczeń, a AI pomaga szybciej rozumieć sytuację. To połączenie może realnie obniżyć ryzyko awarii - pod warunkiem, że zostanie wdrożone z inżynierską ostrożnością, transparentnością i świadomością ograniczeń.

Literatura

Ezgi Gursel, Bhavya Reddy, Anahita Khojandi, Mahboubeh Madadi, Jamie Baalis Coble, Vivek Agarwal, Vaibhav Yadav, Ronald L. Boring, Using artificial intelligence to detect human errors in nuclear power plants: A case in operation and maintenance, Nuclear Engineering and Technology, Volume 55, Issue 2, 2023, Pages 603-622, https://doi.org/10.1016/j.net.2022.10.032.

Ezgi Gursel, Mahboubeh Madadi, Jamie Baalis Coble, Vivek Agarwal, Vaibhav Yadav, Ronald L. Boring, Anahita Khojandi, The role of AI in detecting and mitigating human errors in safety-critical industries: A review, Reliability Engineering & System Safety, Volume 256, 2025, 110682, https://doi.org/10.1016/j.ress.2024.110682.

Esmaeil Zarei, Faisal Khan, Rouzbeh Abbassi, Importance of human reliability in process operation: A critical analysis, Reliability Engineering & System Safety, Volume 211, 2021, 107607, https://doi.org/10.1016/j.ress.2021.107607.

Abu Hawwach M., Human errors in industrial operations and maintenance, Master thesis work, Malardalen University, Sweden, 2021.

M. Blanke, M. Kinnaert, J. Lunze and M. Staroswiecki, Diagnosis and Fault-Tolerant Control, Springer-Verlag, 2004.

J. Korbicz, J.M. Kościelny, Z. Kowalczuk and W. Cholewa (eds), Fault Diagnosis: Models, artificial intelligence methods, applications, Springer, 2004.

R. Isermann, Fault Diagnosis Systems. An Introduction from Fault Detection to Fault Tolerance, Springer, Berlin, 2006.

Weijun Li, Hui Li, Sai Gu, Tao Chen, Process fault diagnosis with model- and knowledge-based approaches: Advances and opportunities, Control Engineering Practice, Volume 105, 2020, 104637, https://doi.org/10.1016/j.conengprac.2020.104637.

Xiaotian Bi, Ruoshi Qin, Deyang Wu, Shaodong Zheng, Jinsong Zhao, One step forward for smart chemical process fault detection and diagnosis, Computers & Chemical Engineering, Volume 164, 2022, 107884, https://doi.org/10.1016/j.compchemeng.2022.107884.

Shuaiqi Yuan, Ming Yang, Genserik Reniers, Integrated process safety and process security risk assessment of industrial cyber-physical systems in chemical plants, Computers in Industry, Volume 155, 2024, 104056, https://doi.org/10.1016/j.compind.2023.104056.

Y. Hu, H. Li, H. Yang et al., Detecting stealthy attacks against industrial control systems based on residual skewness analysis, Journal of Wireless Communications and Networking, 2019, 74, https://doi.org/10.1186/s13638-019-1389-1.

David I. Urbina, Jairo A. Giraldo, Alvaro A. Cardenas, Nils Ole Tippenhauer, Junia Valente, Mustafa Faisal, Justin Ruths, Richard Candell, Henrik Sandberg, Limiting the Impact of Stealthy Attacks on Industrial Control Systems, CCS 2016, Pages 1092-1105, https://doi.org/10.1145/2976749.2978388.

Mica R. Endsley, Supporting Human-AI Teams: Transparency, explainability, and situation awareness, Computers in Human Behavior, 2023.

Mina Saghafian i in., Understanding automation transparency and its adaptive design implications in safety-critical systems, Safety Science, 2025.

Qi Zhang i in., LLM-TSFD: An industrial time series human-in-the-loop fault diagnosis method based on a large language model, Expert Systems with Applications, 2025.

Shuwen Zheng i in., Empirical study on fine-tuning pre-trained large language models for fault diagnosis of complex systems, Reliability Engineering & System Safety, 2024.

Laifa Tao i in., LLM-based framework for bearing fault diagnosis, Mechanical Systems and Signal Processing, 2025.

S. Wen i in., Leveraging large language models for Human-Machine collaborative troubleshooting of complex industrial equipment faults, 2025.

Kaze Du i in., LLM-MANUF: An integrated framework of Fine-Tuning large language models for intelligent decision-making in manufacturing, Advanced Engineering Informatics, 2025.

J. Tang i in., DT and LLM driven intelligent maintenance system, 2025.

Jan Betley i in., Training large language models on narrow tasks can lead to broad misalignment, Nature, 2026.

Xiangyu Qi i in., Fine-tuning Aligned Language Models Compromises Safety, Even When Users Do Not Intend To!, ICLR 2024.

Mrinank Sharma i in., Towards Understanding Sycophancy in Language Models, ICLR 2024.