Wpływ rozwoju modeli AI na dynamikę zagrożeń

Obserwacje oraz zbiór dobrych praktyk przygotowania organizacji

Departament Cyberbezpieczeństwa w UKNF • Warszawa, 23.07.2026 roku

1. Obserwacja ogólna CSIRT KNF

CSIRT KNF[1] obserwuje szybki wzrost zdolności modeli językowych w zakresie programowania, korzystania z narzędzi, pracy agentowej oraz realizacji wieloetapowych zadań technicznych. Zmiana ta jest widoczna zarówno w modelach rozwijanych w Stanach Zjednoczonych, jak i w Chinach, przy czym oba ekosystemy przyjmują częściowo odmienne modele dystrybucji: od usług zamkniętych i dostępu kontrolowanego po publikację otwartych wag umożliwiających lokalne uruchomienie oraz dalsze dostrajanie.

W ocenie CSIRT KNF rozwój ten wpływa na cały cykl życia podatności: zwiększa skalę przeglądu kodu, ułatwia triage i walidację błędów, skraca czas przygotowania proof-of-concept, wspiera analizę poprawek i commitów oraz umożliwia automatyzację rekonesansu i dalszych etapów ataku. Nie oznacza to, że rozwój AI jest jedyną przyczyną wzrostu liczby CVE. Wzrost ten wynika również z większej liczby podmiotów CNA, rozszerzenia praktyk koordynowanego ujawniania, rosnącej złożoności oprogramowania oraz lepszej widoczności błędów.

> **Główna obserwacja**
>
> Znaczenie modeli AI nie polega wyłącznie na tworzeniu nowych klas podatności. Kluczowe jest zwiększenie przepustowości: ta sama liczba ekspertów – lub mniej doświadczony operator wspierany przez agentów – może analizować więcej kodu, szybciej odtwarzać warunki błędu i równolegle testować większą liczbę wariantów ataku.

2. Rozwój modeli w USA i Chinach

Rynek modeli klasy frontier rozwija się równolegle w dwóch głównych ośrodkach. W USA dominują usługi komercyjne z rozbudowanymi mechanizmami bezpieczeństwa i kontrolą dostępu. W Chinach, obok usług API, rozwija się bardzo silny segment open-weight, który umożliwia lokalne uruchamianie modeli oraz ogranicza możliwość monitorowania lub zablokowania nadużyć przez operatora usługi.

[1] Zespół CSIRT KNF realizuje zadania CSIRT sektorowego w rozumieniu ustawy o krajowym systemie cyberbezpieczeństwa.

Ośrodek	Model	Data	Dostęp	Znaczenie z perspektywy podatności
USA	**GPT-5.6 (OpenAI)**[2]	09.07.2026	usługa / API	Zaawansowane kodowanie agentowe, koordynacja narzędzi i wieloagentowość; producent wskazuje istotny wzrost zdolności cyber.
USA	**Claude Fable 5 / Mythos 5 (Anthropic)**[3]	09.06.2026	usługa; dostęp Mythos kontrolowany	Rozdzielenie wersji ogólnej i wersji o ograniczonym dostępie do defensywnych badań cyber pokazuje strategiczną wrażliwość zdolności.
Chiny	**DeepSeek-V4 (DeepSeek)**[4]	24.04.2026	open-weight + API	Otwarta publikacja wariantów Pro i Flash, kontekst 1M oraz silne zdolności agentowe obniżają próg dostępu do zaawansowanego kodowania.
Chiny	**Kimi K3 (Moonshot AI)**[5]	16.07.2026	usługa / API	Model 2,8T i kontekst 1M, ukierunkowany na długie zadania kodowania i pracę agentową; na 21.07 brak publicznej publikacji pełnych wag.
Chiny	**Qwen3.8-Max-Preview (Alibaba)**[6]	19.07.2026	wersja preview w usłudze	Model 2,4T udostępniony w Qoder/Token Plan, rozwijany m.in. pod kątem coding i długich zadań. Nie jest obecnie modelem open-weight.
Chiny	**Qwen3 (Alibaba)**[7]	29.04.2025	open-weight	Istotny kamień milowy otwartego ekosystemu Qwen; publiczne wagi i możliwość lokalnego uruchomienia wielu wariantów.

Tabela 1. Wybrane modele frontier udostępnione w latach 2025–2026 (stan na 21.07.2026).

Uwaga terminologiczna. W debacie publicznej określenie „open source" jest często stosowane „szeroko". W odniesieniu do modeli AI precyzyjniejsze jest rozróżnienie modeli open-weight, dla których opublikowano wagi, od modeli dostępnych wyłącznie w usłudze API lub w wersji preview. Qwen3.8-Max-Preview został udostępniony 19 lipca 2026 roku, ale na dzień opracowania nie jest publikacją otwartych wag. Trend open-weight w ekosystemie chińskim reprezentują m.in. Qwen3 i DeepSeek-V4.

Znaczenie modelu dystrybucji

W przypadku usługi API operator może stosować klasyfikatory bezpieczeństwa, telemetrię i blokady kont. Model open-weight może być uruchomiony lokalnie, dostrojony i pozbawiony zabezpieczeń bez wiedzy

[2] OpenAI – „GPT-5.6: Frontier intelligence that scales with your ambition", 09.07.2026, https://openai.com/index/gpt-5-6/ (dostęp: 23.07.2026).

[3] Anthropic – Claude Fable 5 / Mythos 5 oraz komunikat o ponownym udostępnieniu modeli, czerwiec–lipiec 2026, https://www.anthropic.com/news/redeploying-fable-5 (dostęp: 23.07.2026).

[4] DeepSeek – „DeepSeek-V4 Preview Release", 24.04.2026, https://api-docs.deepseek.com/news/news260424/ (dostęp: 23.07.2026).

[5] Moonshot AI – Kimi K3, 16.07.2026, https://www.moonshot.cn/ (dostęp: 23.07.2026).

[6] Alibaba Cloud – Qwen3.8-Max-Preview, udostępnienie od 19.07.2026, https://help.aliyun.com/zh/lingma/qwen3-8-max-preview-limited-time-offer (dostęp: 23.07.2026).

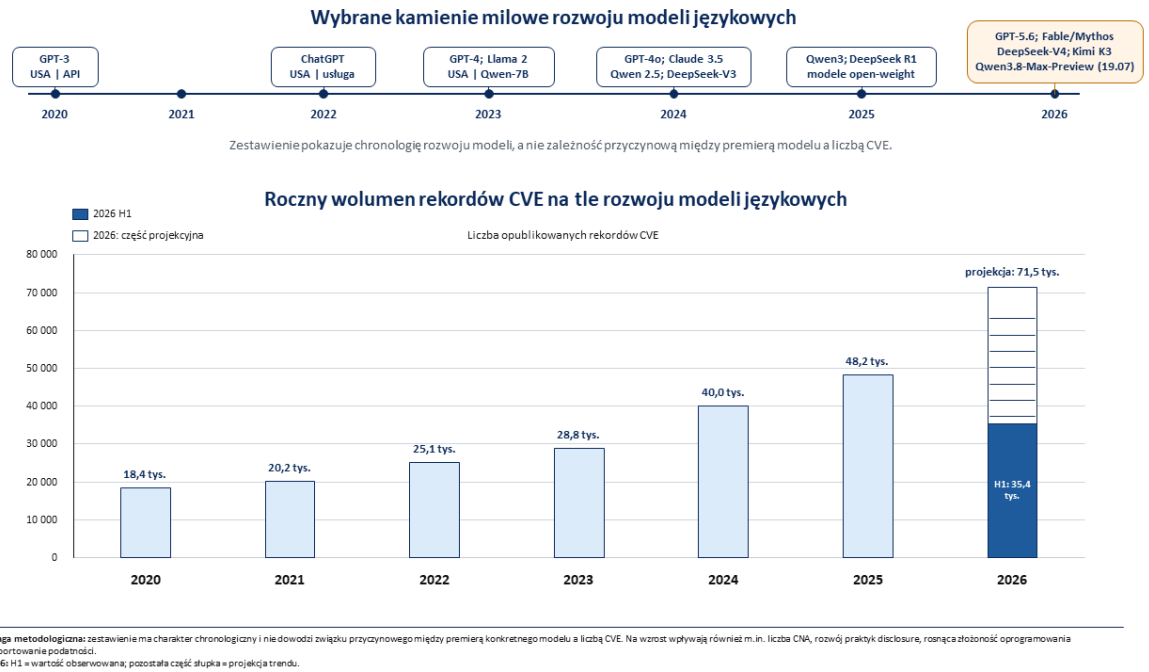
[7] Qwen Team – „Qwen3: Think Deeper, Act Faster", 29.04.2025, https://qwenlm.github.io/blog/qwen3/ (dostęp: 23.07.2026).

pierwotnego dostawcy. Z perspektywy zagrożeń ważna jest więc nie tylko jakość modelu, ale również możliwość jego niekontrolowanego użycia na dużą skalę.

3. Tło statystyczne: CVE, NVD i KEV

CVE List i National Vulnerability Database (NVD)[8] opisują rosnący wolumen ujawnianych podatności, natomiast katalog CISA Known Exploited Vulnerabilities (KEV)[9] wskazuje podzbiór luk, dla których istnieją wiarygodne dowody aktywnego wykorzystania. Te dwa zestawy danych pokazują różne wymiary problemu: presję informacyjną oraz operacyjną.

Liczba publikowanych rekordów CVE wzrosła z ok. 18,4 tys. w 2020 roku do 48,2 tys. w 2025 roku, czyli o ok. 162%. W pierwszym kwartale 2026 roku CVE Program opublikował 15 176 rekordów, o 19% więcej niż w czwartym kwartale 2025 roku[10]. Materiał roboczy wykorzystany do tej analizy wskazuje ok. 35,4 tys. rekordów w pierwszym półroczu 2026 roku oraz scenariusz trendowy rzędu 71–72 tys. w całym roku. Wartość roczna 2026 pozostaje projekcją, nie wynikiem zamkniętym.



Rysunek 1. Roczny wolumen CVE i wybrane kamienie milowe rozwoju modeli[11,12]. Zestawienie chronologiczne nie stanowi dowodu związku przyczynowego. Źródła: CVE Program/NVD oraz oficjalne komunikaty producentów modeli.

[8] NIST – National Vulnerability Database, NVD Dashboard, https://nvd.nist.gov/general/nvd-dashboard (dostęp: 23.07.2026).

[9] CISA – Known Exploited Vulnerabilities Catalog, https://www.cisa.gov/known-exploited-vulnerabilities-catalog (dostęp: 23.07.2026).

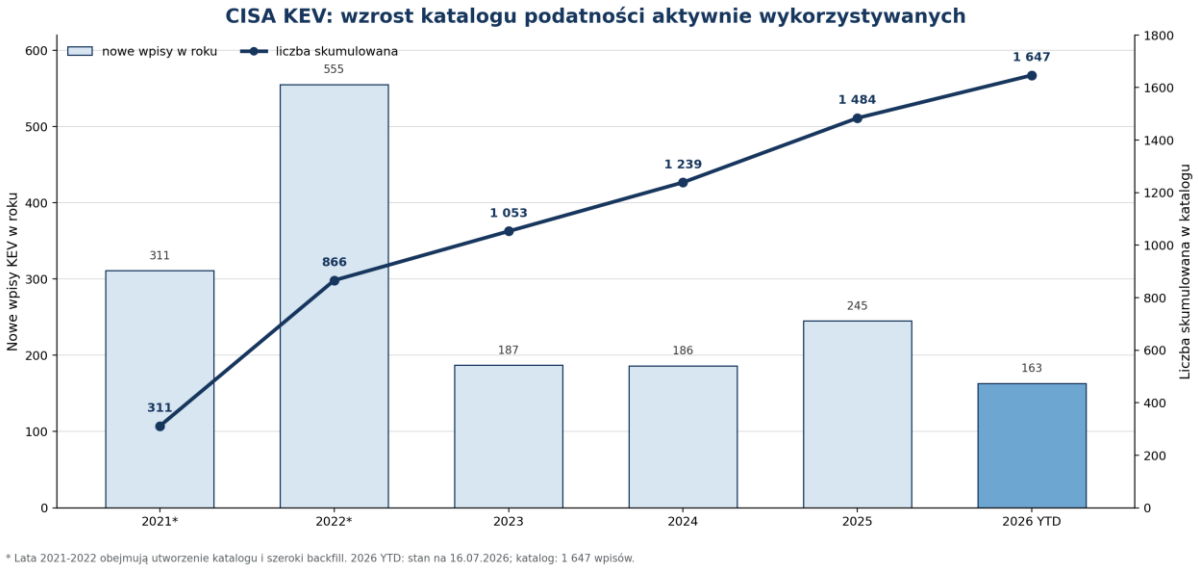
[10] CVE Program – Metrics oraz raport za I kwartał 2026 roku (15 176 opublikowanych rekordów), https://www.cve.org/Media/News/item/blog/2026/05/12/CVE-Program-Report-for-Q1-2026 (dostęp: 23.07.2026).

[11] Qwen Team – „Qwen2.5”, 19.09.2024, https://qwenlm.github.io/blog/qwen2.5/ (dostęp: 23.07.2026).

[12] DeepSeek – „Introducing DeepSeek-V3”, 26.12.2024; „DeepSeek-R1 Release”, 20.01.2025, https://api-docs.deepseek.com/news/news1226 (dostęp: 23.07.2026).

Na wykresie zestawiono dwie równoległe dynamiki: przyrost liczby rekordów CVE oraz kolejne generacje modeli. Zbieżność czasowa jest istotna jako kontekst analityczny, ale nie pozwala przypisać konkretnej części wzrostu CVE do AI. W latach 2023–2026 jednocześnie zwiększała się liczba CNA, dojrzewały procesy disclosure, a część ekosystemów – w szczególności Linux – rozpoczęła szersze i bardziej systematyczne nadawanie identyfikatorów CVE.

Katalog KEV ma bardziej bezpośrednie znaczenie operacyjne, ponieważ obejmuje podatności aktywnie wykorzystywane. Według oficjalnego zbioru CISA katalog liczył 1 647 pozycji w połowie lipca 2026 roku, wobec 1 484 na koniec 2025 roku. Lata 2021–2022 obejmują utworzenie katalogu i szeroki backfill, dlatego ich przyrostów nie należy porównywać wprost z późniejszymi latami.



Rysunek 2. Rozwój katalogu CISA KEV. KEV nie obejmuje wszystkich podatności, lecz luki, dla których potwierdzono aktywną eksploitację. 2026 YTD: stan na 16.07.2026.

> **Wniosek ze statystyk**
>
> Wolumen CVE opisuje rosnącą liczbę informacji wymagających przetworzenia. KEV opisuje rosnącą pulę luk, których wykorzystanie nie jest hipotetyczne. Dla organizacji oznacza to konieczność łączenia źródeł: sam CVSS lub samo NVD nie wystarczają do szybkiej oceny ekspozycji.

4. Wpływ modeli AI na cykl życia podatności

4.1. Odkrywanie i walidacja

Modele połączone z kompilatorami, debuggerami, fuzzerami, repozytoriami kodu i środowiskami testowymi mogą analizować kod w wielu iteracjach, generować przypadki brzegowe, odtwarzać awarie i oceniać, czy błąd jest rzeczywiście wykorzystywalny. Publiczne przykłady obejmują znalezienie przez system Big Sleep nieznanej podatności w SQLite[13] oraz wykrycie przez system MDASH nowych podatności

[13] Google Project Zero – „From Naptime to Big Sleep: Using Large Language Models To Catch Vulnerabilities In Real-World Code”, 01.11.2024, https://projectzero.google/2024/10/from-naptime-to-big-sleep.html (dostęp: 23.07.2026).

w komponentach Windows[14]. Pokazuje to, że AI „przeszła" od streszczania dokumentacji do aktywnego udziału w vulnerability research.

4.2. Skracanie czasu od publikacji do exploita

Po opublikowaniu advisory, commita lub poprawki agent może porównać wersje, wskazać zmieniony fragment, wygenerować przypadki testowe, zbudować proof-of-concept i iteracyjnie poprawiać wynik. Automatyzacja nie musi od razu prowadzić do w pełni autonomicznego ataku. Już system wspierający operatora może znacząco skrócić analizę i umożliwić równoległe testowanie wielu produktów oraz wersji.

4.3. Skalowanie rekonesansu i ekspozycji

Agent AI może zestawiać dane o domenach, adresach IP, certyfikatach, nagłówkach HTTP, endpointach API i wersjach produktów z CVE, KEV, changelogami i wynikami skanerów. Pozwala to przechodzić od ogólnej informacji o podatności do listy potencjalnie narażonych celów. Szczególne znaczenie mają systemy brzegowe, VPN, usługi zdalnego dostępu, aplikacje webowe, API oraz zapomniane zasoby chmurowe.

4.4. Modele open-weight i brak kontroli operatora

Publikacja otwartych wag umożliwia uruchamianie modelu poza infrastrukturą dostawcy, bez telemetrii i bez możliwości zablokowania konta. Model może zostać dostrojony pod zadania ofensywne albo połączony z lokalnymi skanerami, frameworkami eksploitacyjnymi i bazami danych. Wraz ze spadkiem kosztów obliczeniowych zwiększa to liczbę aktorów zdolnych do automatyzacji pracy technicznej.

4.5. Kod generowany przez AI i nowa podaż błędów

Ta sama technologia zwiększa tempo wytwarzania oprogramowania. Kod generowany lub modyfikowany z pomocą AI może zawierać błędy autoryzacji, niewłaściwe zarządzanie sekretami, podatne zależności oraz brak zabezpieczeń charakterystycznych dla dojrzałego procesu SDLC. Ryzyko dotyczy kodu własnego, oprogramowania dostawców, skryptów operacyjnych, low-code/no-code oraz szybkich prototypów przenoszonych do produkcji.

4.6. Przeciążenie procesu remediacji

Zdolność znajdowania i opisywania błędów skaluje się szybciej niż zdolność producentów oraz organizacji do przygotowywania, testowania i wdrażania poprawek. „Wąskim gardłem" stają się: walidacja zgłoszeń, koordynowane ujawnianie, przygotowanie łatki, testy regresji, okna serwisowe oraz zależności dostawców. W praktyce fala podatności może więc przyjąć formę zarówno większej liczby CVE, jak i częstszych pakietów aktualizacji wymagających szybkiej, ale kontrolowanej obsługi.

[14] Microsoft Security Blog – „Defense at AI speed: Microsoft's new multi-model agentic security system tops leading industry benchmark", 12.05.2026, https://www.microsoft.com/en-us/security/blog/2026/05/12/defense-at-ai-speed-microsofts-new-multi-model-agentic-security-system-tops-leading-industry-benchmark/ (dostęp: 23.07.2026).

5. Spójność z „Krajobrazem cyberzagrożeń w polskim sektorze finansowym 2026”

Opisana obserwacja rozwija kierunek wskazany w rozdziale 8 raportu „Krajobraz cyberzagrożeń w polskim sektorze finansowym 2026”[15] przygotowanego przez CSIRT KNF. W raporcie wskazano przejście od AI jako narzędzia wspierającego phishing i generowanie treści do AI jako akceleratora rekonesansu, identyfikacji podatności, przygotowywania wariantów exploitów oraz planowania ataków wieloetapowych.

W rozdziale 8 CSIRT KNF zwraca uwagę, że najbardziej realistycznym scenariuszem nie musi być całkowicie autonomiczny cyberatak. Istotne ryzyko powstaje już wtedy, gdy system AI wspiera operatora: interpretuje wyniki rekonesansu, ustala priorytety, dobiera narzędzia, generuje payloady testowe, porównuje wyniki i wskazuje najbardziej obiecujące ścieżki. Taki model działania zwiększa tempo i skalę operacji, pozostawiając człowieka w roli decydenta.

> **Obserwacja CSIRT KNF zawarta w „Krajobrazie cyberzagrożeń w polskim sektorze finansowym 2026”**
>
> AI może skracać czas potrzebny atakującym na przejście od informacji o podatności do działającego scenariusza ataku. W konsekwencji rośnie znaczenie aktualnego rozpoznania ekspozycji, szybkiej oceny informacji o luce oraz zdolności wdrażania kontrolowanych mitygacji i poprawek.

6. Kontekst europejski: DORA, AI Act i działania organów UE

Zagrożenia opisane w tym opracowaniu są równolegle przedmiotem skoordynowanych działań na poziomie europejskim. Obowiązujące ramy regulacyjne – w szczególności Rozporządzenie DORA oraz AI Act – pozostają podstawą zarządzania tym ryzykiem. DORA zachowuje pełną adekwatność poprzez wymogi dotyczące zarządzania ryzykiem ICT, testowania odporności operacyjnej, zarządzania incydentami oraz ryzykiem ICT stron trzecich, przy zachowaniu zasady proporcjonalności (art. 4 Rozporządzenia DORA): działania powinny uwzględniać wielkość, profil ryzyka oraz charakter, skalę i złożoność działalności podmiotu. AI Act obejmuje dostawców modeli AI ogólnego przeznaczenia, stwarzających ryzyko systemowe dodatkowymi obowiązkami, w tym w zakresie przejrzystości, dokumentacji technicznej i cyberbezpieczeństwa.

W czerwcu i lipcu 2026 roku organy europejskie opublikowały serię skoordynowanych komunikatów. Europejska Rada ds. Ryzyka Systemowego wydała ostrzeżenie w sprawie systemowych ryzyk cybernetycznych związanych z modelami frontier AI, wskazując na możliwość wystąpienia ryzyka jednocześnie w wielu instytucjach poprzez podatności we współdzielonej infrastrukturze i pojedyncze punkty awarii[16]. 7 lipca 2026 roku Komisja Europejska przedstawiła Plan działania na rzecz cyberbezpieczeństwa i sztucznej inteligencji[17], a EBC skierował do prezesów instytucji istotnych pismo

[15] CSIRT KNF – „Krajobraz cyberzagrożeń w polskim sektorze finansowym 2026”, w szczególności rozdział 8, https://www.knf.gov.pl/dla_rynku/CSIRT_KNF/Krajobraz_cyberzagrozen_w_polskim_sektorze_finansowym (dostęp: 23.07.2026).

[16] ERRS (ESRB) – Ostrzeżenie Europejskiej Rady ds. Ryzyka Systemowego z dnia 25 czerwca 2026 r. w sprawie systemowych ryzyk cybernetycznych związanych z modelami frontier AI (ESRB/2026/3), https://www.esrb.europa.eu/pub/pdf/warnings/esrb.warning260625_on_systemic_cyber_risks_stemming_from_frontier_ai_models~ef424708cf.en.pdf (dostęp: 23.07.2026).

[17] Komisja Europejska – „EU Action Plan on Cybersecurity and Artificial Intelligence”, 07.07.2026, https://digital-strategy.ec.europa.eu/en/library/eu-action-plan-cybersecurity-and-artificial-intelligence (dostęp: 23.07.2026).

w sprawie zagrożeń cybernetycznych wspieranych przez AI, oczekując m.in. niezwłocznego zamknięcia otwartych ustaleń nadzorczych w obszarach ICT oraz przedłożenia planów działania do 31 października 2026 roku[18]. Ponadto w lipcu 2026 roku ENISA opublikowała rekomendacje dotyczące rozwoju zdolności operacyjnych w erze frontier AI[19]. Kierunek tych komunikatów jest zbieżny z obserwacją CSIRT KNF: rozwój modeli frontier nie tworzy nowych kategorii ryzyka, lecz istotnie zwiększa tempo i skalę materializacji ryzyk istniejących.

Europejskie podejście nadzorcze porządkuje działania mitygujące w trzech uzupełniających się filarach: prewencji (kompletna inwentaryzacja aktywów wraz z komponentami AI/ML, bezpieczeństwo w fazie projektowania, proaktywny patching, redukcja powierzchni ataku, zarządzanie dostępem i łańcuchem dostaw), detekcji (przejście od okresowego do ciągłego skanowania i monitoringu, w tym behawioralnego, wzmacnianie SOC i red teamingu również z wykorzystaniem AI) oraz zarządzaniu ryzykiem i odpornością operacyjną (testy odporności uwzględniające scenariusze wspierane przez AI i awarie wielosystemowe, plany ciągłości działania, odizolowane kopie zapasowe, odpowiedzialność organu zarządzającego i aktualizacja poziomu tolerancji ryzyka związanego z ICT).

7. Zbiór dobrych praktyk przygotowania organizacji na falę podatności

Poniższy zbiór ma charakter praktyczny i może być wykorzystywany wewnątrz organizacji do oceny gotowości. Nie stanowi katalogu wymagań nadzorczych ani jednolitych terminów dla wszystkich podmiotów. Zakres stosowania powinien uwzględniać krytyczność usług, ekspozycję zasobów oraz możliwości operacyjne organizacji. Praktyki uporządkowano według trzech filarów spójnych z podejściem europejskim: prewencja, detekcja oraz zarządzanie ryzykiem i odporność operacyjna.

Obszar	Dobre praktyki
FILAR I – PREWENCJA	
Widoczność aktywów i ekspozycji	Aktualny rejestr aktywów, wersji, zależności, usług internet-facing, API, urządzeń brzegowych, środowisk chmurowych i zasobów testowych, w tym komponentów AI/ML. Powiązanie aktywów z procesami i właścicielami.
Redukcja powierzchni ataku i ochrona kodu	Eliminacja zbędnych ekspozycji, segmentacja sieci, wycofywanie systemów legacy, zasady secure by design dla systemów krytycznych oraz ochrona poufności kodu źródłowego i konfiguracji przed automatyczną analizą przez modele AI.
Zarządzanie dostępem i tożsamością	Minimalizacja uprawnień (least privilege), MFA odporne na phishing, dostęp just-in-time do systemów krytycznych, monitoring kont uprzywilejowanych oraz relacji zaufania, które mogą być mapowane i wykorzystywane przez agentów AI.
Szybkie, ale kontrolowane aktualizacje	Walidacja źródła i podpisu aktualizacji, staging, testy regresji, plan rollbacku, monitoring po wdrożeniu i jednoznaczna odpowiedzialność za decyzję; docelowo proaktywny, zautomatyzowany patching systemów wysokiego ryzyka, obejmujący również środowiska DR i backup.

[18] EBC – Nadzór Bankowy – pismo do prezesów instytucji istotnych „Addressing AI-enabled cybersecurity threats”, 07.07.2026, https://bankingsupervision.europa.eu/press/letterstobanks/shared/pdf/2026/ssm.2026_letter_on_AI_enabled_cybersecurity_threats.en.pdf (dostęp: 23.07.2026).

[19] ENISA – „ENISA's view on cybersecurity in the frontier AI era”, lipiec 2026, https://www.enisa.europa.eu/publications/enisas-view-on-cybersecurity-in-the-frontier-ai-era (dostęp: 23.07.2026).

Obszar	Dobre praktyki
Kontrole kompensujące	Gotowe scenariusze segmentacji, ograniczenia dostępu, wyłączenia funkcji, WAF/IPS, virtual patching, rotacji sekretów i zwiększonego monitoringu, gdy poprawka nie jest dostępna.
Bezpieczny łańcuch dostaw	SBOM, SCA, pinowanie wersji, kontrola integralności paczek i obrazów, allowlisty komponentów, ograniczanie skryptów instalacyjnych oraz monitoring repozytoriów i CI/CD; egzekwowanie standardów bezpieczeństwa w całym łańcuchu dostaw, w tym u dostawców oprogramowania i sprzętu.
Kod wspomagany AI	Zatwierdzone narzędzia, zakaz przekazywania danych wrażliwych do nieautoryzowanych usług, obowiązkowy review, SAST, secret scanning, testy zależności i adekwatne testy bezpieczeństwa.
FILAR II – DETEKCJA	
Wieloźródłowa informacja o podatnościach	Łączenie CVE/CNA, NVD, KEV, EPSS, advisory producentów, informacji CSIRT oraz własnej telemetrii. Brak pełnego wpisu NVD nie jest traktowany jako dowód braku ryzyka.
Szybki triage	Zdefiniowana ścieżka oceny: czy komponent występuje w organizacji, gdzie jest eksponowany, czy istnieje exploit lub wpis KEV, jaki jest wpływ biznesowy oraz jakie mitygacje są możliwe.
Telemetria i detekcja	Korelacja informacji o podatnościach z logami, EDR/NDR, WAF, IAM i SIEM; monitoring prób wykorzystania, zachowania po aktualizacji oraz anomalii na systemach brzegowych; monitoring behawioralny nastawiony na nietypowe wzorce generowane przez ataki wspierane AI.
Ciągły monitoring i częstsze testy	Przejście od okresowego do ciągłego skanowania podatności, zwiększenie częstotliwości testów penetracyjnych i przeglądów zgodności oraz wzmacnianie zdolności SOC i red teamingu, również z wykorzystaniem rozwiązań AI.
FILAR III – ZARZĄDZANIE RYZYKIEM I ODPORNOŚĆ OPERACYJNA	
Ład i odpowiedzialność zarządu	Jednoznaczna odpowiedzialność organu zarządzającego za ryzyko związane z modelami frontier AI, przejście od okresowego przeglądu do bieżącego nadzoru, aktualizacja poziomu tolerancji ryzyka związanego z ICT (w tym metryki, progi tolerancji) oraz adekwatne zasoby do wzmacniania cyberodporności.
Ciągłość działania i backupy	Plany reagowania na incydenty i plany ciągłości działania uwzględniające awarie wielosystemowe i równoległą eksploitację wielu podatności; kopie zapasowe odizolowane od ryzyk środowiska produkcyjnego i regularnie testowane odtwarzanie.
Testy odporności operacyjnej	Okresowe testy odporności rozwijane w kierunku scenariuszy ataków wspieranych przez AI, w tym awarii kaskadowych w systemach współzależnych; skalowalne, oparte na analizie zagrożeń podejście do zakresu testów.
Współpraca z dostawcami	Aktualne kontakty eskalacyjne, dostęp do advisory, informacje o wersjach i zależnościach, zasady przekazywania SBOM oraz uzgodniony sposób działania przy podatności bez dostępnej poprawki.
Świadomość i kompetencje	Programy edukacyjne zwiększające świadomość dot. zagrożeń, uwzględniające socjotechnikę wspieraną przez AI (spersonalizowany phishing, deepfake); wzmacnianie zdolności pracowników do wykrywania i zgłaszania anomalii jako uzupełniającego elementu detekcji.
Ćwiczenie „fala podatności”	Scenariusz obejmujący publikację krytycznej luki w „szeroko używanym: komponencie, ustalenie ekspozycji, wdrożenie mitygacji, ocenę skuteczności poprawki, komunikację i monitoring prób wykorzystania.

Tabela 2. Dobre praktyki przygotowania organizacji na falę podatności, uporządkowane według filarów: prewencja – detekcja – zarządzanie ryzykiem.

> **Ważne!**
> Model docelowy nie oznacza bezrefleksyjnego wdrażania każdej aktualizacji natychmiast, ale zdolność do szybkiego ustalenia ekspozycji i przeprowadzenia kontrolowanego, weryfikowalnego procesu: mitygacja, test, aktualizacja, rollback i monitoring.

8. Wnioski

Rozwój modeli w USA i Chinach prowadzi do wzrostu dostępności zdolności programistycznych, agentowych i cyber. W USA najsilniejsze funkcje są na ogół dostarczane w usługach kontrolowanych, natomiast chiński ekosystem równolegle rozwija usługi frontier i modele open-weight. To drugie zjawisko ma szczególne znaczenie, ponieważ umożliwia lokalne uruchomienie oraz modyfikację modeli poza kontrolą operatora.

Statystyki CVE i KEV pokazują, że organizacje już funkcjonują w warunkach rosnącego wolumenu informacji o podatnościach oraz stale powiększającej się puli luk aktywnie wykorzystywanych. AI nie jest jedyną przyczyną tego wzrostu, ale staje się istotnym akceleratorem wyszukiwania, walidacji i wykorzystania błędów. Najważniejszą konsekwencją jest skrócenie czasu dostępnego na ocenę i działanie.

Z perspektywy CSIRT KNF właściwą odpowiedzią jest przygotowanie organizacji na częstsze fale ujawnień i poprawek: lepsza widoczność aktywów, wieloźródłowy monitoring, szybki triage, kontrolowane aktualizacje, gotowe mitygacje, bezpieczeństwo łańcucha dostaw oraz uwzględnienie kodu wspomaganego AI w procesach bezpieczeństwa oprogramowania. Kierunek ten jest spójny z działaniami organów europejskich oraz z wytycznymi Rozporządzenia DORA i AI Act, w których dodatkowo podkreślono rolę organu zarządzającego, ciągłość monitoringu oraz odporność operacyjną na awarie wielosystemowe.

9. Uwagi metodologiczne

Dane z 2026 roku mają charakter bieżący i mogą zmieniać się wraz z kolejnymi publikacjami CVE, aktualizacjami katalogu KEV oraz zmianami dostępności modeli. Projekcja CVE na cały 2026 rok została wyraźnie oznaczona. Zestawienie premier modeli na wykresie służy przedstawieniu kontekstu czasowego i nie jest analizą ekonometryczną ani dowodem przyczynowości. Źródła wykorzystane w opracowaniu wskazano w przypisach dolnych, w miejscu ich przywołania.