

„Scenariusze jutra”

Krótki opis wyzwania : Światowa polityka zmienia się coraz szybciej, a czynników ją kształtujących jest coraz więcej. W tym chaosie informacyjnym musimy znaleźć prawidłowości i trendy. Do tego potrzebne jest nam narzędzie, dzięki któremu będziemy mogli działać wyprzedzająco i odkodowywać symptomy przyszłych wydarzeń już dziś!

1. Wprowadzenie - opis organizacji, sytuacji i stanu aktualnego

Na czym polega praca w MSZ? W dużej mierze na operowaniu informacjami – znajdowaniu ich, przetwarzaniu i podejmowaniu na ich podstawie decyzji. Setki pracowników MSZ – w kraju i na placówka zagranicą – spędzają rocznie dziesiątki tysięcy godzin na pracy z informacjami, ustalaniu związków przyczynowo-skutkowych, wskazywaniu możliwych scenariuszy na przyszłość i rekomendowaniu rozwiązań. To zadanie czasochłonne, a przy tym obarczone ryzykiem pominięcia ważnych informacji, co może negatywnie wpływać na wyciąganie wniosków i formułowanie rekomendacji.

Tu ważną rolę może odegrać tzw. analiza foresightowa – uporządkowany proces przetwarzania informacji nt. zdarzeń obecnych i przeszłych, który ma prowadzić do tworzenia możliwych i racjonalnych scenariuszy sytuacji międzynarodowej na przyszłość. Obecnie dysponujemy pewnymi narzędziami do analizy danych i prognozowania wydarzeń geopolitycznych, które jednak wciąż są stosunkowo mało wyrafinowane. Podstawowym podmiotem analitycznym jest wciąż ludzki umysł. Stąd ograniczone moce analizowania tysięcy stron danych i wyekstrahowania z nich elementów, które mogą być symptomami przyszłych wydarzeń. Żadna instytucja nie może sobie przecież pozwolić, aby nad jednym zadaniem pracowało 100 czy 1000 „Kowalskich”. Dotychczasowe narzędzia – i ludzie w szczególności – umożliwiają analizę ograniczonej liczby parametrów bazując na dość prostych założeniach. Wpływa to negatywnie na zdolność do przewidywania długoterminowych scenariuszy politycznych, gospodarczych i obronnych. Mimo postępów w uczeniu maszynowym i analizie big data, wciąż brakuje dostosowanych do potrzeb dyplomacji narzędzi, które byłyby w stanie uwzględnić interakcje wieloczynnikowe i wielostronne czy zmienne społeczne.

W efekcie prognozy często mają charakter uproszczony i bywają podatne na błędy, nie dostarczając wyczerpujących wyjaśnień przyczyn obserwowanych trendów i ich konsekwencji. Brakuje systemów, które mogłyby automatycznie zintegrować dziesiątki czy nawet setki parametrów i zależności, aby generować elastyczne, wielowarstwowe scenariusze (np. na 12 lub 36 miesięcy). Dostępne obecnie narzędzia analityczne zazwyczaj nie tłumaczą powiązań między różnymi danymi wejściowymi. Bez tego samo proponowanie prognoz jest obarczone dużym ryzykiem niskiej wiarygodności i ograniczonej użyteczności dla osób podejmujących strategiczne decyzje.

Celem wyzwania jest zatem stworzenie praktycznego narzędzia, które wyjdzie poza obecne ograniczenia – będzie skalowalne i zaprojektowane w taki sposób, by w przyszłości analizować nawet 50- lub 100-krotnie większą liczbę parametrów (w stosunku do prototypu opracowanego podczas hackathonu). Dzięki temu pracownicy MSZ będą mogli nie tylko lepiej

rozumieć aktualne trendy, ale również przygotowywać wielopoziomowe scenariusze międzynarodowe, uwzględniające szeroki wachlarz czynników politycznych, gospodarczych, społecznych i obronnych. Zbudowanie takiego narzędzia ma na celu zwiększenie zdolności do wiarygodnego wskazywania przyszłych trendów i adaptacji w zmiennym, niepewnym środowisku międzynarodowym. Wiemy, że dyplomaci różnych krajów już zaczynają korzystać z takich narzędzi. Jeśli my tego nie będziemy robić – zostaniemy z tyłu.

2. Wyzwanie

Cel wyzwania:

Stwórz prototyp skalowalnego narzędzia, które wykorzystuje zaawansowane technologie NLP, analizy danych (deep research) oraz modelowania scenariuszy do typowania prawdopodobnych wydarzeń i trendów w polityce międzynarodowej. Zaprojektowane narzędzie zostanie przetestowane na przykładzie wyimaginowanego państwa członkowskiego UE i NATO o nazwie Atlantis.

Nazwa państwa: Atlantis

Istotne cechy położenia geograficznego: dostęp do Morza Bałtyckiego, kilka dużych żeglowanych rzek, ograniczone zasoby wody pitnej

Liczba ludności: 28 mln

Klimat: umiarkowany

Silne strony gospodarki: przemysł ciężki, motoryzacyjny, spożywczy, chemiczny, ICT, ambicje odgrywania istotnej roli w zakresie OZE, przetwarzania surowców krytycznych oraz budowy ponadnarodowej infrastruktury AI (m.in. big data centers, giga fabryki AI, komputery kwantowe)

Liczebność armii: 150 tys. zawodowych żołnierzy

Stopień cyfryzacji społeczeństwa: powyżej średniej europejskiej

Waluta: inna niż euro

Kluczowe relacje dwustronne: Niemcy, Francja, Finlandia, Ukraina, USA, Japonia

Potencjalne zagrożenia polityczne i gospodarcze: niestabilność w UE, rozpad UE na grupy „różnych prędkości” pod względem tempa rozwoju oraz zainteresowania głębszą integracją; negatywna kampania wizerunkowa ze strony kilku aktorów państwowych wymierzona przeciw rządowi lub społeczeństwu Atlantis; zakłócenia w dostawach paliw węglowodorowych z USA, Skandynawii, Zatoki Perskiej (wynikające z potencjalnych zmian w polityce wewnętrznej krajów eksporterów lub problemów w transporcie, np. ataki Hutich na gazowce na Morzu Czerwonym); narażenie na spowolnienie rozwoju sektora ICT z powodu embarga na wysokozaawansowane procesory

Potencjalne zagrożenie militarne: zagrożenie atakiem zbrojnym jednego z sąsiadów; trwające od wielu lat ataki hybrydowe co najmniej jednego sąsiada, w tym w obszarze infrastruktury krytycznej i cyberprzestrzeni

Kamienie milowe w rozwoju politycznym i gospodarczym: demokracja parlamentarna od 130 lat; okres stagnacji gospodarczej w latach 1930-1950 oraz 1980-1990; członkostwo w UE i NATO od roku 1997; 25. gospodarka świata wg PKB od roku 2020; deficyt budżetowy oraz dług publiczny w okolicach średniej unijnej

Cecha krytyczna narzędzia

Z punktu widzenia użytkownika narzędzia (pracownika MSZ) kluczowa jest wyjaśnialność generowanych informacji. Narzędzie musi umieć przedstawić logikę swojego działania (ścieżka przechodzenia od faktów historycznych do faktów spodziewanych, opis zidentyfikowanych korelacji między zbiorami faktów historycznych oraz przyszłych) oraz identyfikować nieoczywiste czynniki zmian na podstawie analizy dużych wolumenów tekstu. Czyli krótko mówiąc – **chain of thought**.

3. Oczekiwany rezultat

Rozwiązańiem problemu (jakim jest zbyt niska w stosunku do potrzeb wydajność ludzkiego umysłu w analizowaniu informacji i typowaniu przyszłych scenariuszy) będzie stworzenie narzędzia do analizy danych tekstowych i liczbowych oferującego jako efekt analizy wieloaspektowe scenariusze zdarzeń przyszłych o realistycznym poziomie probabilizmu.

Narzędzie ma mieć postać aplikacji / agenta, który po określaniu przez urzędnika danych wejściowych (w obecnym wyzwaniu: opisane w punkcie 7 + ew. pliki dostarczone przez MSZ), ma przeanalizować wskazane dane i wyciągnąć z nich wnioski wg kryteriów i wag określonych przez użytkownika. Na potrzeby niniejszego wyzwania jedynym kryterium jest „interes państwa Atlantis” (czyli narzędzie powinno traktować wnioski dotyczące np. trzech małych wysp na Pacyfiku powinno traktować jako nieistotne). Natomiast wagi przedstawione są poniżej a-f.

Używając modeli LLM należy tak ustawić „temperaturę”, aby generowane wyniki były raczej realistyczne niż będące przejawem rozbudowanej fantazji. Jeśli czas pozwoli, zainteresowane zespoły mogą jako „entertaining exercise” wygenerować jeden wybrany scenariusz z „temperaturą” ustawioną bardziej w kierunku fantazji i kreatywności – z zachowaniem zasad(chain of thought).

Gotowe narzędzie zbudowane w ramach obecnego wyzwania ma wygenerować scenariusze (predykcje) oraz rekomendacje dotyczące wpływu sytuacji międzynarodowej na interesy państwa Atlantis. Wyobraźmy sobie, że ich wykonanie zleca (= wykonuje prompt) ambasador (nazywany oficjalnie stałym przedstawicielem) państwa Atlantis przy Unii Europejskiej. Materiał ma być przydatny dla rządu państwa Atlantis **w dwóch perspektywach czasowych (12 i 36 miesięcy)** oraz **w dwóch wariantach rozwoju: negatywnym i pozytywnym** dla interesów państwa Atlantis. Zatem łącznie powinny to być cztery scenariusze.

Narzędzie ma wygenerować wspomniane wyżej scenariusze w oparciu o następujący opis sytuacji międzynarodowej:

- a) Wskutek zaistniałej przed miesiącem katastrofy naturalnej wiodący światowy producent procesorów graficznych stracił 60% zdolności produkcyjnych; odbudowa

- mocy produkcyjnych poprzez inwestycje w filie zlokalizowane na obszarach nieobjętych katastrofą potrwa do końca roku 2028 (waga istotności: 30)
- b) Przemysł motoryzacyjny w Europie (piętnaście głównych partnerów handlowy państwa Atlantis to kraje europejskie) bardzo wolno przedstawia się na produkcję samochodów elektrycznych; rynek europejski zalewają tanie samochody elektryczne z Azji Wschodniej; europejski przemysł motoryzacyjny będzie miał w roku 2025 zyski na poziomie 30% średnich rocznych zysków z lat 2020-2024 (waga istotności: 15)
 - c) PKB krajów strefy euro w roku 2025 spadnie średnio o 1,5% w stosunku do roku 2024 (waga istotności: 15)
 - d) Na wschodzie Ukrainy trwa słaby rozejm; Rosja kontroluje dwie główne elektrownie ukraińskie, które pracują na potrzeby konsumentów rosyjskich; gospodarka ukraińska rozwija się w tempie 4% PKB, głównie dzięki inwestycjom w przemysł zbrojeniowy i odbudowę infrastruktury (waga istotności: 10)
 - e) Inwestycje amerykańskie w Ukrainie kierowane są do przemysłu wydobywczego (surowce krytyczne); roczne inwestycje UE w Ukraine są na poziomie 3% ukraińskiego PKB i utrzymują się na takim poziomie do roku 2029 (waga istotności: 5)
 - f) Mamy gwałtowny wzrost udziału energii z OZE w miesieci energetycznym krajów UE oraz Chin od początku roku 2028; w połowie roku 2023 średnio wielkości kraj południowoamerykański odkrył ogromne i łatwe do eksploatacji złoża ropy naftowej i gazu ziemnego dorównujące wielkością złożom Arabii Saudyjskiej i Kataru, co przełoży się pod koniec roku 2027 na nadpodaż tych paliw na światowe rynki; wzrost podaży energii z OZE oraz nadpodaż paliw węglowodorowych przekładają się na znaczny spadek cen ropy: do poziomu 30-35 USD za baryłkę; będzie to miało wpływ na budżet Rosji oraz (w mniejszym stopniu) innych krajów producentów ropy i paliw ropopochodnych (waga istotności: 25)

Materiał końcowy ma mieć postać tekstu (może być z danymi liczbowymi i wykresami graficznymi) o objętości 2-3 tysięcy słów. Tekst powinien mieć następującą strukturę:

- a) Streszczenie danych uwzględnionych w analizie (max. 250 słów) – w sposób przejrzysty, user-friendly
- b) Przedstawienie poszczególnych scenariuszy (po każdym scenariuszu powinna znaleźć się część wyjaśniająca korelacje między poszczególnymi elementami analizowanych informacji oraz wskazująca na związki przyczynowo-skutkowe między informacjami a wnioskami, na bazie których powstały scenariusze)
- c) Rekomendacje dla państwa Atlantis:
 - Jakie decyzje pomogą uniknąć scenariuszy negatywnych
 - Jakie decyzje pomogą w zrealizowaniu scenariuszy pozytywnych

4. Wymagania formalne

Projekt przesyłany do oceny powinien zawierać:

- szczegółowy opis i tytuł projektu,

- film umieszczony w dostępnym, otwartym repozytorium (link), trwający maksymalnie 3 minuty i prezentujący projekt (działanie narzędzia od momentu określenia danych wejściowych po uzyskanie tekstu zgodnego z opisem w punkcie 3).
- Opis architektury rozwiązania
- Tekst zgodny z opisem w punkcie 3
- Optymalnie: narzędzie w wersji demo

Dodatkowo może zawierać:

- repozytorium kodu,
 - materiały graficzne lub inne elementy związane z projektem
-

5. Wymagania techniczne

MSZ co do zasady nie ogranicza wolności wyboru zespołom kodującym co do wykorzystywanych narzędzi, jednak optymalne będzie użycie języka Python. Narzędzie może zostać zbudowane w formule uwzględniającej dostęp zespołów kodujących do sieci internetowej ORAZ korzystanie z plików dostarczonych przez MSZ.

Kwestie cyberbezpieczeństwa powinny być traktowane priorytetowo przy konstrukcji narzędzia. Żaden użytkownik sieci internetowej niezalogowany do domeny MSZ nie może mieć wglądu do promptów wysyłanych do chmur obliczeniowych oraz do stworzonych przez narzędzie tekstów analitycznych. Dotyczy to także poszczególnych zapytań (których mogą być tysiące) składających się na pojedynczy prompt. Właścicielem promptów może być tylko urzędnik MSZ i MSZ jako instytucja.

Do stworzenia narzędzia kodujący mogą korzystać wyłącznie z licencji bezpłatnych.

Narzędzie musi umożliwiać przetwarzanie dużych zbiorów danych tekstowych w języku polskim i angielskim. Wersja podstawowa musi analizować do 50 mln słów (oficjalne komunikaty ministerstw, wybranych think-tanków, raporty giełd, raporty dot. rynków surowcowych, rezolucje ONZ/EU/NATO, itp.). Powinno posiadać opcję pamięci 10 ostatnich promptów i ich rezultatów – jeśli nie w wersji przedstawionej do oceny, to przynajmniej z możliwością rozwoju w tym kierunku.

6. Sposób testowania i/lub walidacji

Zapoznanie się z funkcjonowaniem narzędzia (prezentacja i/lub demo) oraz produktami jego działania (tekst analityczno-prognostyczny). Oceniający sprawdzają jakość doboru analizowanych danych przez narzędzie, jakość i wyjaśnialność wnioskowania i

dostosowanie scenariusza oraz ew. rekomendacji do interesów państwa Atlantis. Ocenie podlegać też będzie intuicyjność obsługi interface'u narzędzia.

7. Dostępne zasoby

Do właściwego wykonania zadania konieczne będzie korzystanie przez budowane narzędzie z następujących stron internetowych:

- anglojęzyczne strony ministerstw spraw zagranicznych, ministerstw obrony, ministerstw spraw wewnętrznych, ministerstw gospodarki, ministerstw handlu, ministerstw energii, ministerstw klimatu, ministerstw szkolnictwa wyższego, ministerstw nowych technologii / cyfryzacji, ministerstw edukacji następujących krajów: Niemcy, Francja, UK, USA, Rosja, Chiny, Indie, Arabia Saudyjska;
 - anglojęzyczne strony następujących instytucji: Komisja Europejska, NATO, ONZ, OECD, Gulf Cooperation Council, International Institute for Strategic Studies, Center for Strategic and International Studies, Chatham House, European Council on Foreign Relations, Atlantic Council, Kiel Institute, NASDAQ, London Stock Exchange Group, Japan Exchange Group). W ramach zleconego zadania narzędzie ma analizować dane publikowane po 31 grudnia 2020 r.
-

8. Kryteria oceny

- Związek z wyzwaniem — 30%
- Pomyśl — 10%
- Oryginalność — 5%
- Kompletność projektu — 15%
- Potencjał wdrożeniowy — 40%

Kluczowym aspektem oceny narzędzia jest przejrzystość jego działania i wyjaśnialność generowanych wyników.

Każda prognoza na przyszłość musi opisywać logiczną ścieżkę przyczynową: mechanika priorytetyzacji faktów / danych wpływających na predykcję, sposoby rozwiązywania kontradykcji faktycznych lub interpretacyjnych w analizowanych materiałach. Przy czym sam użytkownik musi mieć także możliwość „ręcznego” określania wag poszczególnych elementów danych (a nawet wydania instrukcji, by narzędzie skorygowało swój produkt zgodnie z wprowadzonymi przez urzędnika zmianami wag czy parametrów).

9. Dodatkowe uwagi / kontekst wdrożeniowy

Jeśli przedstawione narzędzie zostanie ocenione jako przydatne dla MSZ, posiadające potencjał rozwoju funkcjonalności oraz zbudowane w sposób umożliwiający zaimplementowanie elementów gwarantujących wymagany przez MSZ poziom ochrony przetwarzanych informacji, zwycięskiemu zespołowi może zostać zaproponować dalsza współpraca.

Wersja podstawowa (oddana do oceny) musi umożliwiać gruntowną rozbudowę:

- wzrost zdolności obliczeniowej: analiza do 5 mld słów w pojedynczej operacji wywołanej promptem (czyli 100x więcej niż wersja podstawowa)
- możliwość rozszerzenia parametrów geograficznych do 50 krajów / 30 języków (do 30x więcej niż w wersji podstawowej)
- możliwość zawężania danych wejściowych na podstawie takich parametrów, jak rejestracja domeny publikującej dane w krajach o określonej liczbie ludności, określonych zasobach wody, określonej liczbie dni nasłonecznienia lub dni wietrznych, posiadających/nieposiadających dostęp do morza, posiadających/nieposiadających broń jądrową, posiadających/nieposiadających zasoby paliw kopalnych
- możliwość rozszerzenia poddawanych analizie materiałów poza dane tekstowe - o dane graficzne, audio i wideo (m.in. w formatach jpg, gif, tif, pdf, docx, txt, xml, mp3, mp4, wave, cdr, ai, psd)
- możliwość dodawania nowych parametrów (innych niż np. geograficzne, językowe, dot. formatów danych)
- możliwość pracy narzędzia na danych zamkniętych (kontenery) – z odłączeniem od sieci publicznej
- wielowariantowość metody analitycznej - zmiana konfiguracji działania narzędzia skutkująca zmianą metody analitycznej, np. poprzez zastosowanie prognozowania wstecznego (backcasting) polegającego na określeniu założonego stanu przyszłego, a następnie cofaniu się krok po kroku w czasie aż do stanu obecnego, dzięki czemu urzędnik będzie mógł uzyskać wiedzę o etapach koniecznych do zaistnienia pożdanego (lub wręcz przeciwnie) stanu przyszłego.

Narzędzie powinno posiadać warstwę (np. oddzielna zakładka), która przedstawi ścieżkę „rozumowania” prowadzącego do poszczególnych wniosków. Poszczególne jednostki informacji (np. cena ropy w dacie X lub nałożenie embarga na produkty Y przez kraj Z) powinny być otagowane. Narzędzie powinno w tej warstwie wskazywać, w jaki sposób poszczególne informacje prowadzą do konkretnych wniosków i które informacje traktowane są priorytetowo (i dlaczego). W tej warstwie powinna być także wskazana ścieżka prowadząca z poszczególnych wniosków lub grup wniosków do rekomendacji.

Bonus dla MSZ

Ogromną zaletą narzędzia będzie zaimplementowanie mechanizmu uodporniającego narzędzie na „data poisoning”, czyli celowe zanieczyszczanie ogólnodostępnych danych przez nieprzychylnych aktorów, co ma skutkować znaczącym obniżeniem jakości generowanych wyjściowych materiałów analitycznych.

10. Kontakt

Nazwa stoiska: Ministerstwo Spraw Zagranicznych - “Scenariusze jutra”

Sebastian Rejak sebastian.rejak@msz.gov.pl

Alexander Nowakowski alexander.nowakowski@msz.gov.pl
