

CZY MOŻNA PRZEWIDYWAĆ KRYZYSY WIZERUNKOWE W ORGANIZACJACH Z POMOCĄ SZTUCZNEJ INTELIGENCJI ?

WYNIKI BADAŃ ZREALIZOWANYCH W RAMACH PROJEKTU CRISIS DETECTOR

dr hab. inż. Grzegorz Chodak

Politechnika Wrocławska, Katedra Sztucznej Inteligencji

prof. Dariusz Tworzydło

Uniwersytet Warszawski, Katedra Komunikacji Społecznej
i Public Relations

Konsultacje: Kaja Matuszczak

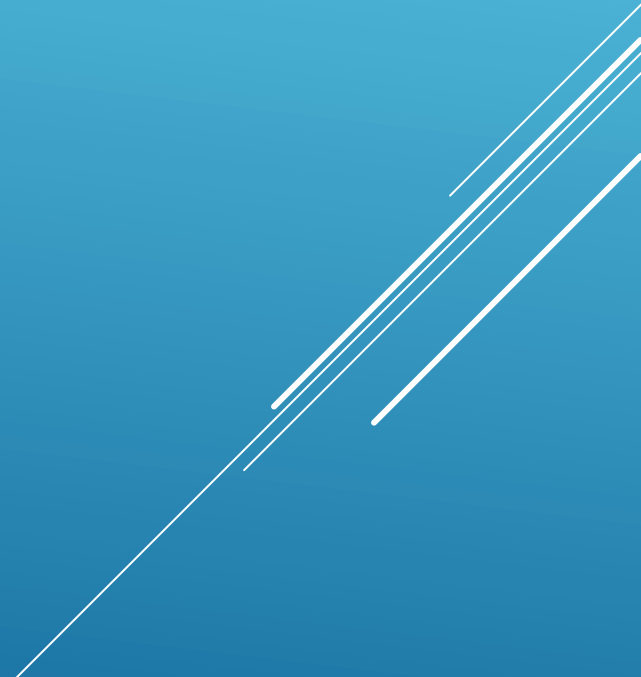
Instytut Monitorowania Mediów



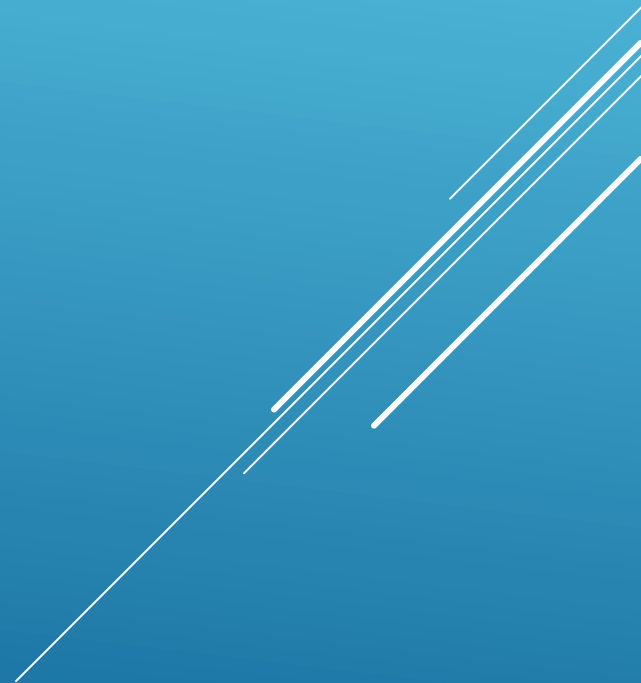
PLAN

1. Kilka słów o projekcie Crisis Detector
2. Kryzysy wizerunkowe
3. Zaproponowana klasyfikacja kryzysów
4. Wyniki badań

PROJEKT CRISIS DETECTOR

- ▶ Współpraca z Instytutem Monitorowania Mediów
 - ▶ Koordynator: prof. Maciej Piasecki
 - ▶ Analiza strumienia danych medialnych
 - ▶ Realizacja w latach 2023-2024
 - ▶ Publikacja wyników 2025
- 
- Several white lines of varying lengths and angles are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

KRYZYSY WIZERUNKOWE



BADANE MEDIA

- ▶ Analiza strumieni tekstu wszystkich dostępnych liczących się mediów (baza Inforia):
 - ▶ 350 stacji TV/Radio
 - ▶ TV - TVP, TVN, Polsat, Trwam, ...
 - ▶ Radio - RMF FM, Radio ZET, Polskie Radio,...
 - ▶ 1700 tytułów prasowych - Gazeta Wyborcza, Gość Niedzielny, Rzeczpospolita, Fakt, Newsweek Polska, ...
 - ▶ 21 tysięcy portali internetowych (Onet, WP, Interia...)
 - ▶ Media społecznościowe (Facebook, Twitter/X, YouTube, and Instagram, Wykop, Reddit)

PROCES PRZYGOTOWYWANIA DANYCH – ETAP WSTĘPNY

- ▶ Analiza tekstów medialnych w kontekście zbioru słów kluczowych dotyczących kryzysów (lata 2016 - 2023)
- ▶ Wyodrębnienie 300 kryzysów wizerunkowych organizacji i osób publicznych
 - ▶ Ręczna weryfikacja kryzysów
 - ▶ Dostępność dużej liczby doniesień medialnych – wybór na podstawie fraz charakterystycznych dla danego kryzysu

PRZYKŁADOWE KRYZYSY WIZERUNKOWE PODDAWANE ANALIZIE

- ▶ Kaufland - dyskryminacja płacowa kobiet / według wyników kontroli PIP w sieci dochodziło do dyskryminacji pracowników, korzystających z uprawnień rodzicielskich.
- ▶ McDonald's - użycie w reklamie WieśMaka motywów z Symfonii Leningradzkiej Szostakowicza
- ▶ Leroy Merlin - bojkot marki związany z jej pozostaniem na rosyjskim rynku
- ▶ Uniwersytet Gdański - na ulotkach, rozdawanych na organizowanych przez uczelnię Dniach Mózgu, znalazły się hasła, promujące (bezpieczne) zażywanie narkotyków
- ▶ Mateusz Morawiecki - afera premiowa po Mistrzostwach Świata w Katarze
- ▶ Radosław Sikorski i wątek Zjednoczonych Emiratów Arabskich

ANALIZA KRYZYSÓW

- ▶ Określenie punktu wystąpienia kryzysu
- ▶ Analiza dwa miesiące wstecz i miesiąc później

USTALENIE GROUND TRUTH

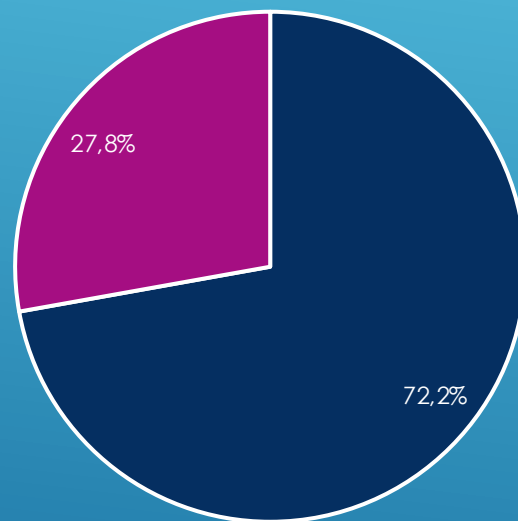
1. Anotacja tekstów (100 tekstów na kryzys)

1. Czy kryzys występuje
2. Anotacje emocyjne (wyniki nie zostały opublikowane)

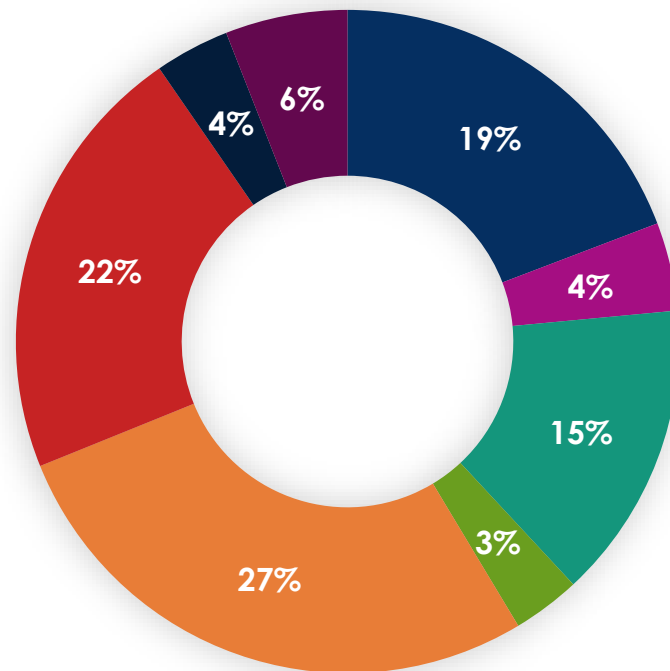
2. Typologia 300 kryzysów

- ▶ Dwie anotacje eksperckie
- ▶ Wysoka zgodność (PSA ponad 0.9)
- ▶ Wspólne ustalenie ground truth

Kryzysy dotyczące organizacji lub jej członków a kryzysy dotyczące osób publicznych / celebrytów



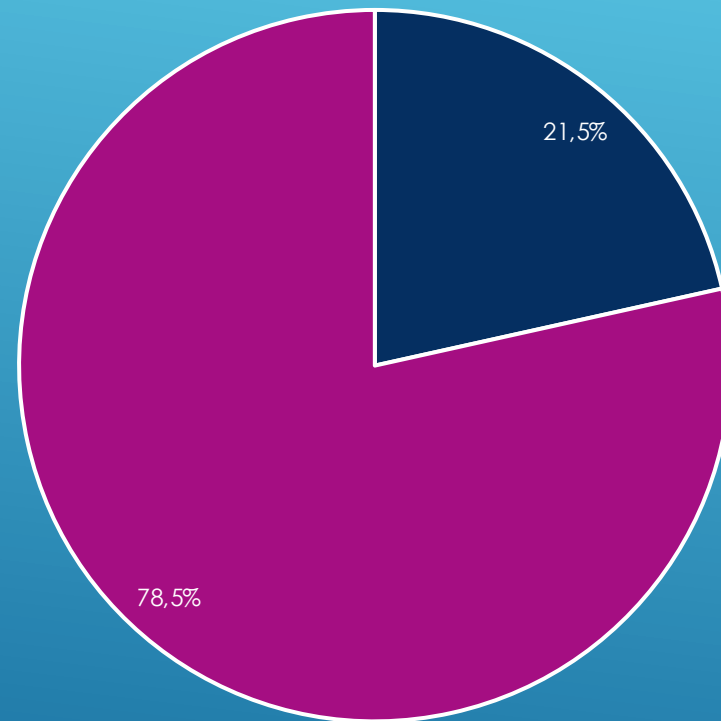
- Kryzysy dotyczące organizacji lub jej członków
- Kryzysy dotyczące osób publicznych / celebrytów



- błędy organizacji
- działania/zachowania celebryty/osoby publicznej nieakceptowalne społecznie
- kontrowersyjna kampania reklamowa
- kontrowersyjna wypowiedź członka organizacji
- kontrowersyjne działania organizacji lub działania/zachowania jej członka
- kontrowersyjne działania/zachowania celebryty/osoby publicznej
- nieakceptowalne społecznie działania organizacji lub działania/zachowania jej członka
- wypowiedź celebryty/osoby publicznej

KLASYFIKACJA KRYZYSÓW ZE WZGLĘDU NA PRZYCZYNY

Kryzysy ideologiczne/nieideologiczne



■ Kryzys ideologiczny ■ Kryzys nie-ideologiczny

IDENTYFIKACJA KRYZYSU Z WYKORZYSTANIEM LLMA - EKSPERYMENTY

- ▶ Przygotowanie pobudzeń
- ▶ LangChain pipeline
- ▶ Testowane modeli:
 - ▶ ChatGPT-3.5-turbo (pierwsze badania)
 - ▶ ChatGPT-4.0-turbo (pierwsze badania)
 - ▶ ChatGPT-4o-mini
 - ▶ ChatGPT-5
 - ▶ ChatGPT-5-mini
 - ▶ o4-mini

TREŚĆ POBUDZENIA

Jako system wczesnego ostrzegania przed kryzysami wizerunkowymi, Twoim zadaniem jest monitorowanie internetu w celu wykrywania tekstów i komentarzy dotyczących naszych klientów, którzy to są firmami i znanymi osobami.

Twoim zadaniem jest ostrzeganie nas, gdy rozpoczyna się kryzys wizerunkowy dotyczący któregoś z nich.

Otrzymasz NAZWA_WĄTKU oraz OPIS, na podstawie których będziesz musiał określić typ kryzysu oraz podmiot, którego on dotyczy.

TREŚĆ POBUDZENIA – OKREŚLENIE CZY KRYZYS MA PODŁOŻE IDEOLOGICZNE

Określ typ i podmiot kryzysu na podstawie podanej poniżej

NAZWA_WĄTKU

i OPISU.

NAZWA_WĄTKU:``{thread\_name}``OPIS:``{description}``

Typ kryzysu określamy w dwóch wymiarach: pierwszy dotyczy obecności podłoża ideologicznego, drugi – przyczyn kryzysu.

Twoja odpowiedź powinna być przedstawiona w formacie JSON, zgodnie z poniższym wzorem:{"ideology": int, "cause": int, "entity": str}}

Wartość pola "ideology" powinna wynosić 1, jeśli kryzys ma podłoże ideologiczne, i 0 w przeciwnym przypadku.

Podłoże ideologiczne może być związane z jednym z następujących poglądów: konserwatyzm, lewicowość, liberalizm, socjalizm, komunizm, anarchizm, nacjonalizm, faszyzm, feminizm, ekologizm, genderyzm, ideologia LGBT, ideologia antyszczepionkowa, ideologia pandemiczna/antypandemiczna, ideologia antymigracyjna, ideologia migracyjna, ideologia antyreligijna, ideologia proaborcyjna.

TREŚĆ POBUDZENIA – OKREŚLENIE PRZYCZYNY KRYZYSU ZE WZGLĘDU NA JEJ PRZYCZYNĘ

Wartość pola "cause" powinna odpowiadać poniższej specyfikacji:

- 0: Jeśli przyczyną kryzysu jest błąd organizacji.
- 1: Jeśli przyczyną kryzysu są działania konkurencji oraz inne zewnętrzne ataki lub działania.
- 2: Jeśli przyczyną kryzysu jest kontrowersyjna wypowiedź członka organizacji.
- 3: Jeśli przyczyną kryzysu jest kontrowersyjna kampania reklamowa.
- 4: Jeśli przyczyną kryzysu są kontrowersyjne działania organizacji lub działania/zachowania jej członka.
- 5: Jeśli przyczyną kryzysu są nieakceptowalne społecznie działania organizacji lub działania/zachowania jej członka.
- 6: Jeśli przyczyną kryzysu jest wypowiedź celebryty/osoby publicznej.
- 7: Jeśli przyczyną kryzysu są kontrowersyjne działania/zachowania celebryty/osoby publicznej.
- 8: Jeśli przyczyną kryzysu są działania/zachowania celebryty/osoby publicznej nieakceptowalne społecznie.

Wartość pola "entity" powinna odpowiadać osobie lub organizacji, której dotyczy kryzys, będącej w ogniu krytyki.

TREŚĆ POBUDZENIA – FEW SHOT

W swoich odpowiedziach weź pod uwagę podane przykładowe kryzysy - pole WYNIK zawiera poprawną anotację, która została przygotowana przez specjalistę.

Przykłady:

NAZWA WĄTKU `texttt{[ORGANIZATION]}` - seksistowski wywiad z `texttt{[PERSON_NAME]}`

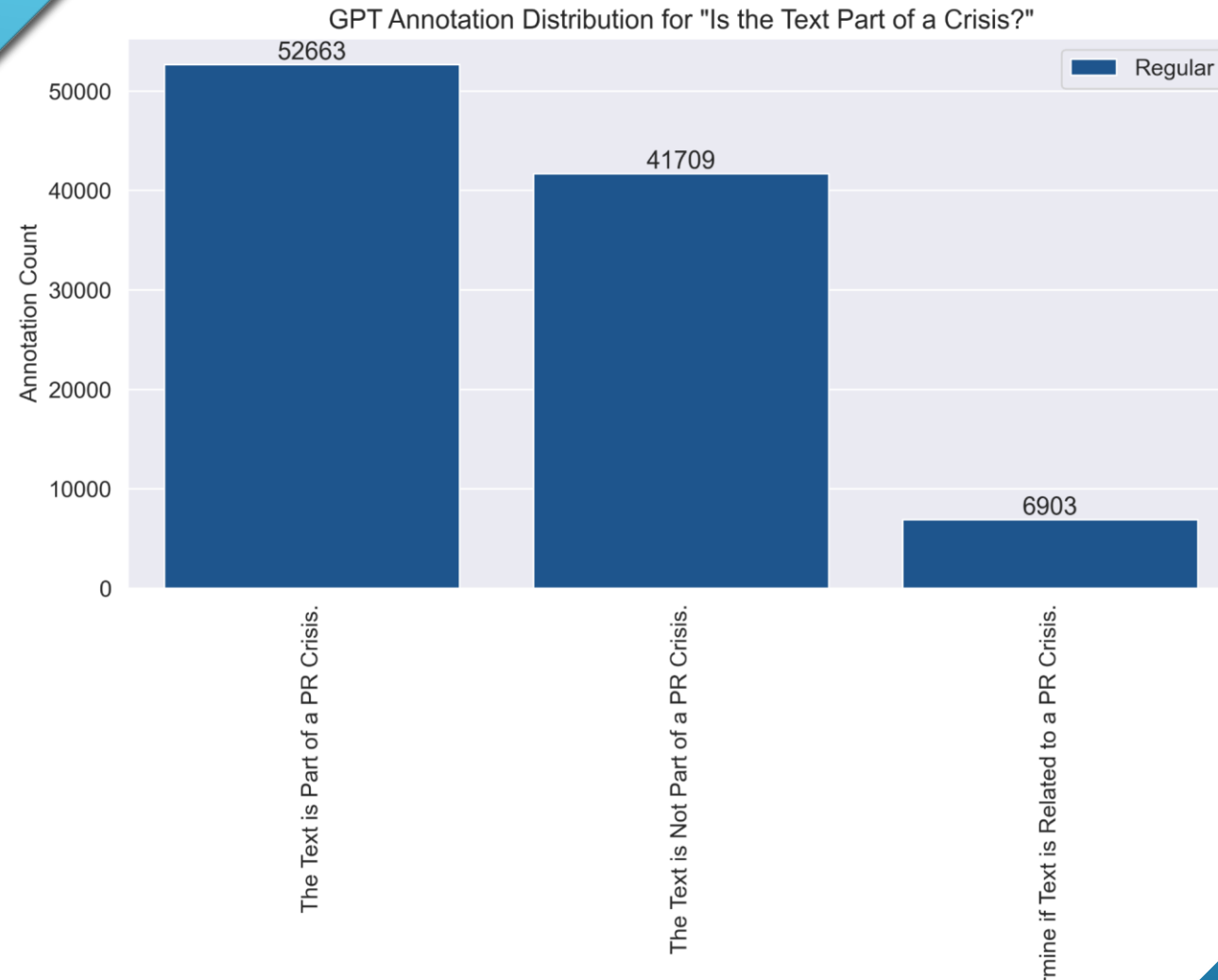
OPIS: `texttt{[ORGANIZATION]}` opublikował wywiad, w którym `texttt{[PERSON_NAME]}` zadaje kontrowersyjne pytania aktorce `texttt{[PERSON_NAME]}` - odtwórczyni głównej roli w filmie "365 dni" była pytana m.in. o to, czy lubi seks i czy założyła na casting koronkową bieliznę. Na publikację krytycznie zareagowało stowarzyszenie Kobiet Filmu.

WYNIK: `{{'ideology': 1, 'cause': 2, 'entity': [ORGANIZATION] / [PERSON_NAME]}}`

WYNIKI

- F1-Score
 - Sugerowana miara ze względu na ryzyko niezbalansowanych klas
- Accuracy
 - Będzie zawyżona dla niezbalansowanych danych
- Ewaluacja wyników – ground truth na podstawie anotacji dwóch ekspertów

ROZKŁAD KLASYFIKACJI PRZEZ MODEL GPT



- ▶ Tekst zawiera informacje o kryzysie
- ▶ Tekst nie zawiera informacji o kryzysie
- ▶ Model nie potrafił stwierdzić, czy tekst zawiera informacje o kryzysie

WYNIKI BADAŃ – ACCURACY

Model	Entity affected by the crisis	Typology by cause of the crisis	Typology of crisis (ideological / non-ideological)	Mean (across tasks)
GPT-3.5-turbo	0.839	0.468	0.758	0.688
GPT-4-turbo	0.952	0.581	0.903	0.812
GPT-4o-mini	0.984	0.516	0.710	0.737
GPT-5	0.968	0.629	0.871	0.823
GPT-5-mini	0.935	0.597	0.790	0.774
o4-mini	0.935	0.581	0.871	0.796
Baseline	N/A	0.306	0.758	0.532
Random Forest	N/A	0.226	0.758	0.492
SVM	N/A	0.290	0.758	0.524
XGBoost	N/A	0.290	0.726	0.508

WYNIKI BADAŃ – F1-SCORE

Model	Entity affected by the crisis	Typology by cause of the crisis	Typology of crisis (ideological/ non-ideolog.)	Mean (across tasks)	
GPT-3.5-turbo	0.839	0.392	0.662	0.631	
GPT-4-turbo	0.952	0.591	0.874	0.806	
GPT-4o-mini	0.984	0.492	0.676	0.717	
GPT-5	0.968	0.597	0.843	0.803	
GPT-5-mini	0.935	0.565	0.757	0.752	
o4-mini	0.935	0.534	0.832	0.767	
Baseline	N/A	0.052	0.431	0.242	
Random Forest	N/A	0.126	0.431	0.279	
SVM	N/A	0.050	0.431	0.240	
XGBoost	N/A	0.155	0.546	0.351	

F1-SCORE DLA PODEJŚCIA 1-SHOT, 2-SHOT

Model	Entity affected by the crisis	Typology by cause of the crisis	Typology of crisis (ideological / non-ideological)	Mean (across tasks)
o4-mini 2 samples	0.968	0.612	0.814	0.798
o4-mini 1 sample	0.968	0.612	0.814	0.798
GPT-5-mini 2 samples	0.952	0.676	0.480	0.703
GPT-5-mini 1 sample	0.968	0.664	0.757	0.796
GPT-5 2 samples	0.984	0.726	0.827	0.846
GPT-5 1 sample	0.968	0.678	0.861	0.836
GPT-4o-mini 2 samples	0.968	0.511	0.620	0.700
GPT-4o-mini 1 sample	0.935	0.524	0.634	0.698
GPT-4-turbo 2 samples	0.952	0.526	0.849	0.776
GPT-4-turbo 1 sample	0.968	0.515	0.824	0.769
GPT-3.5-turbo 2 samples	0.935	0.294	0.741	0.657
GPT-3.5-turbo 1 sample	0.952	0.321	0.719	0.664

PUBLIKACJA



ORIGINAL ARTICLE

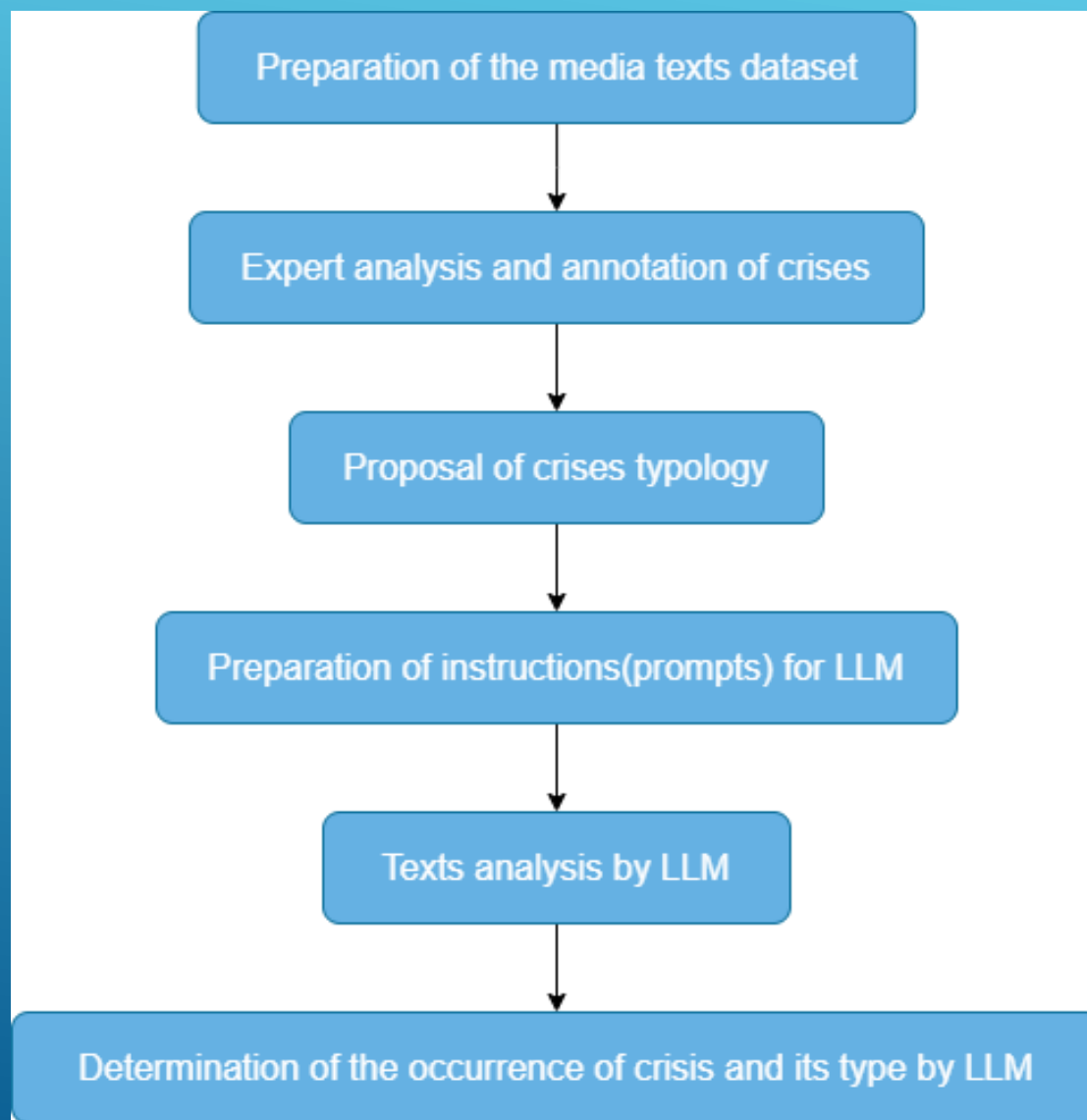
Typology of Image Crises Using Large Language Models: A Novel Approach to Crisis Classification

Chodak, G., Tworzydło, D., Szczesny, A., Kazienko, P., Kaszyca, O., Bilski, K., Oleksy, M., Kochanek, M., Szydło, D., Cichecki, I., Matuszak, K., Mieleszczenko-Kowszewicz, W., Dziecioł, E., Palacz, P., Kajdanowicz, T., Piasecki, M. & Kocoń, J. (2025).

Typology of Image Crises Using Large Language Models: A Novel Approach to Crisis Classification.

Journal of Contingencies and Crisis Management, 33(4), e70092.

PODSUMOWANIE



WNIOSKI

- ▶ LLMy bardzo dobrze sobie radziły z zadaniem określania czy mamy do czynienia z kryzysem, nawet w podejściu 0-shot
 - ▶ Binarna klasyfikacja kryzysów była zadaniem, z którym LLMy poradziły sobie dość dobrze
 - ▶ Bardziej złożona klasyfikacja, wymagała zastosowania podejścia few-shot
- 
- Several white lines of varying lengths and angles are drawn on the right side of the slide, extending from the middle to the bottom right corner.