

# Wstęp

**Temat:** Wykorzystanie GenAI w jakościowej analizie danych

**Długość trwania:** 8h (1 dzień roboczy)

- **Start:** 9:00
- **Koniec:** 17:00

**Wymagania sprzętowe:**

1. Dostęp do Google Colab
2. Dostęp do modelu / modeli GenAI poprzez API:
  - a. API Key
  - b. Endpoint URL

Kwestie do rozwiązania:

1. Czy mają dostęp do Google Colab?
2. Czy mają dostęp do jakiś modeli przez API? Do jakich?

**Wymagania po stronie prowadzącego:**

1. Testowa baza danych
2. Skrypt w google colab
3. Możliwość podłączenia się do Techlandowego API

## Agenda

1. Przedstawienie się: (do 20min)
  - a. Imię i Nazwisko
  - b. Rola, obowiązki, doświadczenie w analizie danych jakościowych
  - c. Ulubiony / najbardziej użyteczny / najzabawniejszy przypadek wykorzystanie GenAI w życiu lub pracy
2. Kontrakt warsztatowy (do 15min)
3. Jakie postawy i sposoby myślenia sprzyjają wykorzystaniu GenAI? (do 35min)
  - a. Dyskusja: (do 20min)
    - i. Z jakich GenAI korzystacie w pracy?
    - ii. Do jakich zadań korzystacie z GenAI a do jakich nie korzystacie?

- iii. W jaki sposób podchodzicie do ulepszania / optymalizowania waszego wykorzystania GenAI?
  - b. Prezka+Q&A: Moja perspektywa na (współ)pracę z GenAI (do 15min)
- 4. Przerwa 15min ~10:30
- 5. Wprowadzenie do problemu analizy danych jakościowych: (do 30min)
  - a. Dotychczasowe problemy z analizą danych jakościowych
  - b. Eksploracja vs. Klasyfikacja
  - c. Uzasadnienie wyboru problemu klasyfikacji na start
  - d. Dlaczego korzystamy z GenAI via API
    - i. Możliwość ewaluacji i optymalizacji
- 6. Zapoznanie się z wyzwaniem i danymi warsztatowymi + Q&A (do 30min)
  - a. Co będziemy robić: klasyfikacja komentarzy ze Steam do The Beast
  - b. Jak wygląda Colab - nie boimy się programować
  - c. Zapoznanie się z przygotowanymi kategoriami do klasyfikacji i ich definicjami
  - d. Jak wygląda zestaw danych ewaluacyjnych
  - e. Co ta ewaluacja w naszym przypadku tak dokładnie znaczy
- 7. Wytlumaczenie zasad konkursu (gamification)
- 8. **Pierwsza Iteracja Promptowania:** Spróbuj sam! (do 35min)
  - a. Wytlumaczenie zadania (5min)
  - b. Praktyka (max. 30 min)
- 9. Przerwa obiadowa + kawka 45-60min ~13:45/14:00
- 10. **Druga Iteracja Promptowania:** Structured Output! (do 45min)
  - a. Dobre praktyki formatowania promptów i dlaczego to ma sens
    - i. XML / md
  - b. Struktura zawartości promptu
  - c. Czym jest Structured Output?
  - d. Jak wykorzystać Pydantic?
  - e. Praktyka (max. 30min)
- 11. **Trzecia Iteracja Promptowania:** Chain-of-Thought (do 45min)
  - a. Czym jest CoT?
  - b. Wady i zalety CoT
  - c. Praktyka (max. 30min)
- 12. Przerwa 15min
- 13. **Czwarta Iteracja Promptowania:** Zero, One, Few, Shoot! (do 45min)
  - a. O co chodzi w Zero, One i Few Shootach?
  - b. Przykłady pozytywne i negatywne
  - c. Praktyka (max. 30min)
- 14. Wyłonienie zwycięzcy + podsumowanie + Q&A (do 30min) ~16:45