

# SYSTEM ANONIMIZACJI TEKSTU

Automatyczna detekcja i anonimizacja danych wrażliwych  
w polskich tekstach konwersacyjnych

Rozmowa z klientem: **Jan Kowalski**, numer **PESEL  
90120512345**, adres **ul. Warszawska 10, 00-001**  
Warszawa. Proszę o pilny kontakt w sprawie  
konta bankowego **1234 5678 9012 3456**.

Rozmowa z klientem: {imię}, numer  
**PESEL** {pesel}, adres {adres},  
Proszę o pilny kontakt w sprawie  
konta bankowego {numer konta}.



Oparty na bibliotece spaCy (Entity Rulers + Regex)



# Architektura Systemu



Detekcja za pomocą spaCy + Reguły i wyrażenia regularne → Proces anonimizacji

- Wykorzystanie modeli językowych dla języka polskiego (np. pl\_core\_news\_lg)
- Entity Rulers do definiowania niestandardowych reguł dopasowania
- Wyrażenia regularne do wykrywania danych o określonym formacie (PESEL, telefony, email)



# Kluczowe Funkcje



## Świadomość Kontekstu

Rozróżnianie podobnych typów danych na podstawie otaczającego tekstu (np. 'Pan Dąbrowski' vs 'miejscowość Dąbrowa')



## Obsługa Tekstów Nieformalnych

Efektywne przetwarzanie dialogów, slangu i tekstów z błędami typowymi dla komunikacji nieformalnej



## Generowanie Danych Syntetycznych

Tworzenie realistycznych, fałszywych danych zachowujących format i charakterystykę oryginałów

**Przed:** "Dzwonił Jan Kowalski z numeru 555-123-456 w sprawie konta 1234 5678 9012 3456"

**Po:** "Dzwonił {imię\_1} {nazwisko\_1} z numeru {telefon\_1} w sprawie konta {nr\_karty\_1}"

# Zastosowania i Korzyści



## Ochrona Prywatności w Biznesie

- Anonimizacja transkrypcji rozmów z call center
- Bezpieczne przechowywanie komunikacji z klientami
- Ochrona danych w dokumentach firmowych



## Wsparcie Badań Naukowych

- Bezpieczne wykorzystanie danych medycznych
- Analiza sentymetru i zachowań użytkowników
- Szkolenie modeli językowych (LLM)



## Zgodność z Regulacjami

- Spełnianie wymogów RODO (GDPR)
- Minimalizacja ryzyka naruszeń prywatności
- Automatyzacja procesów ochrony danych

System wspiera organizacje w efektywnym zarządzaniu danymi przy jednoczesnej ochronie prywatności

# Podsumowanie



**System Anonimizacji Tekstu to:**

- Zaawansowane rozwiązanie do automatycznej detekcji i anonimizacji danych wrażliwych
- Połączenie biblioteki **spaCy** z **Entity Rulers** i **wyrażeniami regularnymi**
- Narzędzie świadome kontekstu z obsługą tekstu nieformalnych
- Wsparcie dla ochrony prywatności, badań i zgodności z RODO

Skuteczna ochrona prywatności przy  
zachowaniu użyteczności danych

