

# Klasyfikator naturalności stylu tekstu – autentyczny czy syntetyczny

Wykonawcy: Natalia Rucińska, Karol Baran, Jan Tobolewski, Tomasz Truszkowski

Sztuczna Inteligencja, Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska

Opiekun projektu: dr hab. inż. Piotr Szczuko, prof. uczelni

## Dziennikarz czy bot?

Czy teksty na portalach społecznościowych przekazują prawdziwe informacje?

Artykuł odnosi sukces gdy wzbudza duże emocje, porusza tematy takie jak polityka, porady około-medyczne lub teorie spiskowe. Takie teksty można generować przez boty - programy do generowania fałszywych informacji (ang. fake news).

Celem projektu jest przygotowanie klasyfikatora naturalności stylu tekstu – syntetyczny czy autentyczny. Zbiorem danych wejściowych będą artykuły medyczne – z powodu wysokiej szkodliwości nieprawdziwych informacji.

## Przykłady nienaturalnego stylu

*"If you have a positive test you should be able to get an infection test done," said Dr. Michael K. Korn, chief of the Department of Infectious Diseases at the University of Washington. "If you have a negative test you should be able to get an infection test done."*

*"The best way to stop this disease is to stop it from spreading, and to prevent it from spreading to others," said lead author and virologist Dr. Michael S. S. Sargent of Vanderbilt University in Nashville, Tennessee. "It's not a good idea to try to get vaccinated."*

## Klasyfikacja

Zastosowanie grafowych sieci neuronowych, które:

- Przetwarzają nieuporządkowane dane wejściowe;
- Uwzględniają sąsiedztwo elementów (czyli słów), z którego wynikają charakterystyczne cechy, wpływające m.in. na sens, "sentymencie", znaczenie i styl;
- Tekst będzie zatem grafem, klasyfikowanym jako autentyczny lub syntetyczny;

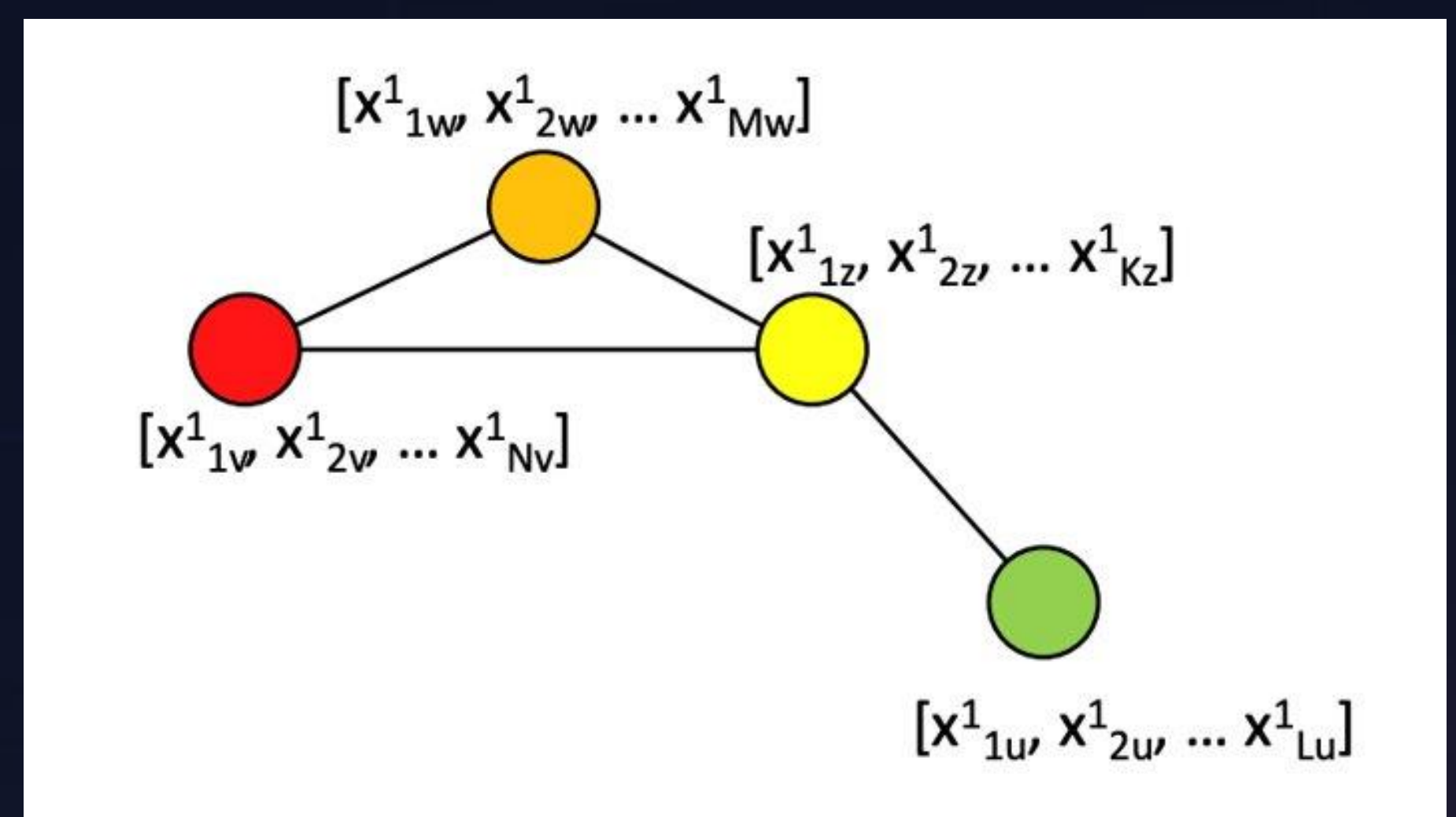


Rys. 1. Fake news w Internecie  
źródło : flaticon.com

## Generowanie tekstów

NLP (ang. Natural Language Processing) to przygotowanie testów tak, aby były zrozumiałe dla maszyn, co umożliwia głębszą analizę całych zbiorów danych.

Artykuł „Attention is all you need” zaproponował architekturę Transformera, na której oparto model GPT-2 (ang. Generative Pre-trained Transformer 2). W projekcie wykorzystamy możliwość dotrenowania modelu do generowania tekstów z określonej dziedziny. Modyfikacje parametrów temperatury i top\_k umożliwi uzyskanie zadowalających tekstów zbioru danych treningowych.



Rys. 2. Głębokie przetwarzanie tekstu i sygnału mowy  
źródło: wykład 12. Grafowe sieci neuronowe; dr hab. inż. Piotr Szczuko