

## **Przetwarzanie języka naturalnego i modelowanie języka**

Rafał Jaworski, specjalność Sztuczna Inteligencja

### **1. Charakterystyka obszaru badawczego**

W mojej pracy zajmuję się zagadnieniami analizy i przetwarzania języka naturalnego, w szczególności języka pisanego, z wykorzystaniem nowoczesnych metod sztucznej inteligencji oraz uczenia maszynowego. Badania te obejmują zarówno modelowanie struktury języka na poziomie leksykalnym, składniowym i semantycznym, jak i automatyczne wydobywanie informacji z dużych zbiorów tekstowych. Szczególną uwagę poświęcam zagadnieniom takim jak reprezentacja znaczenia, analiza kontekstu, rozpoznawanie zależności między elementami tekstu oraz ocena jakości i spójności generowanych treści.

Moja działalność badawcza koncentruje się również na praktycznych zastosowaniach metod NLP, w tym na automatyzacji analizy dokumentów, wspomaganii wyszukiwania informacji oraz budowie systemów wspierających interakcję człowiek-komputer za pomocą języka naturalnego. Istotnym elementem pracy jest także analiza ograniczeń i wyzwań współczesnych modeli językowych, takich jak ich interpretowalność, odporność na błędy oraz wpływ jakości danych treningowych na wyniki. Takie podejście pozwala łączyć perspektywę teoretyczną z aplikacyjną, a jednocześnie przyczynia się do odpowiedzialnego i świadomego wykorzystania technologii przetwarzania języka naturalnego.

### **2. Motywacja**

Przetwarzanie języka naturalnego (NLP) przy użyciu sztucznej inteligencji stanowi interdyscyplinarny obszar badawczy łączący informatykę, lingwistykę oraz statystykę. Obszar ten rozwija się bardzo dynamicznie, a jego znaczenie rośnie wraz z zapotrzebowaniem na automatyzację analizy informacji tekstowych w biznesie, nauce i administracji.

### **3. Obecny poziom badań i możliwości finansowania**

Prowadzone przeze mnie badania mają charakter badań własnych, nie są finansowane przez instytucje zewnętrzne.

### **4. Tematyka badawcza**

- Analiza sentymentu
- Tłumaczenie automatyczne
- Budowanie zasobów lingwistycznych (terminologia, korpusy)
- LLM – eliminacja halucynacji, wyjaśnialna sztuczna inteligencja
- Języki regionalne – analiza i generowanie wypowiedzi

### **5. Wymagania odnośnie członków projektu**

- Znajomość technik programowania w Python
- Zrozumienie architektury sieci neuronowych

### **6. Literatura**

- Attention Is All You Need – Vaswani et al. (2017)
- Improving Language Understanding by Generative Pre-Training – Radford et al. (2018)
- Language Models are Few-Shot Learners – Brown et al. (2020)
- Pre-training of Deep Bidirectional Transformers for Language Understanding – Devlin et al. (2018)
- A Comprehensive Overview of Large Language Models – Naveed et al. (2023)