Doctoral School of Information and Biomedical Technologies
Polish Academy of Sciences (TIB PAN)

SUBJECT:

Large language models supervision powered by common knowledge datasets

SUPERVISOR:

Mariusz Kamola, NASK PIB, Mariusz.Kamola@nask.pl

DESCRIPTION:

LLM output explainability is considered a necessary and sometimes a sufficient condition in the discourse towards safe and ethical AI. The proposed topic is supposed to encompass a range of research activities with aim to develop algorithms that map text documents representations at various stages of DNN processing onto RDF knowledge graphs. Such mappings would in practice take a form of specialized neural networks, trained on commonly available “bilingual” corpora such as DBpedia. The successful mapping functions are enablers of LLM output veracity and coherence verification – either directly or with support of classical automated reasoners.

BIBLIOGRAPHY:

- Futia, G., & Vetrò, A. (2020). On the integration of knowledge graphs into deep learning models for a more comprehensible AI—Three challenges for future research. Information, 11(2), 122.
- Qi, Z., & Li, F. (2017, December). Learning explainable embeddings for deep networks. In NIPS Workshop on Interpretable Machine Learning (Vol. 31).
- Qureshi, M. A., & Greene, D. (2019). EVE: explainable vector based embedding technique using Wikipedia. Journal of Intelligent Information Systems, 53, 137-165.