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

SUBJECT:

Fast and reliable data completion and augmentation with matrix and tensor methods.

SUPERVISOR:

dr hab. inż Mariusz Kamola (main supervisor);
mariusz.kamola@nask.pl;
Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy (NASK-PIB),
ul. Kolska 12, Warszawa

dr Antonina Krajewska (assistant supervisor)
antonina.krajewska@nask.pl
Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy (NASK-PIB),
ul. Kolska 12, Warszawa

DESCRIPTION:

In today's data-driven era, the reliability of machine learning models depends heavily on the quality of datasets. However, real-world data often suffer from missing or noisy values due to various factors like sensor errors, privacy concerns, and the high cost associated with data collection. There is significant interest in developing robust methods for completing matrices and, more recently, tensors to address the missing data problem. The performance and correctness of these algorithms depend on assumptions regarding the properties of the matrices or tensors and the distribution of known elements within them.

The project focuses on developing matrix and tensor completion methods to address these challenges with reference to NASK projects, particularly in network traffic reconstruction and medical image processing. Specific objectives include:

- Research on algorithms of compressed sensing, matrix completion, and tensor completion.
- Ensuring alignment with NASK activities.
- Addressing challenges in incomplete datasets in problems like network traffic reconstruction and medical image processing, or equivalent.
- Validating effectiveness through experimentation.

The research methodology involves literature review, algorithm development, implementation, experimentation, and refinement. Expected outcomes include novel algorithms suitable for NASK projects, deepened theoretical understanding of the subject and dissemination through publications and presentations.

REQUIREMENTS:

- MSc degree in mathematics, computer science or a related field,
- High programming skills in at least one of Python or Matlab
- Advanced Level in English (speaking and writing).

BIBLIOGRAPHY:

[1] Chen, Y., & Chi, Y. (2018). Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization. *IEEE Signal Processing Magazine*, *35*(4), 14-31.

[2]Qingquan Song, Hancheng Ge, James Caverlee, and Xia Hu. 2019. Tensor Completion Algorithms in Big Data Analytics. ACM Trans. Knowl. Discov. Data 13, 1, Article 6 (February 2019), 48 pages. https://doi.org/10.1145/3278607

[3] Liu, A., & Moitra, A. (2020). Tensor completion made practical. Advances in Neural Information Processing Systems, 33, 18905-18916.

[4] Kumar, Awnish, Vijaya V. Saradhi, and Venkatesh Tamarapalli. "Compressive sensing of Internet traffic data using relative-error bound tensor-CUR decomposition." *Journal of Network and Computer Applications* 224 (2024): 103830.

[5] Moitra A. Algorithmic Aspects of Machine Learning. Cambridge University Press; 2018.