GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Explainable AI , PG_00053343
Field of study	Biomedical Engineering, Biomedical Engineering, Biomedical Engineering

Date of commencement of studies	February 2024	Academic year of realisation of subject	2024/2025
Education level	second-cycle studies	Subject group	Optional subject group Subject group related to scientific research in the field of study
Mode of study	Full-time studies	Mode of delivery	at the university
Year of study	2	Language of instruction	Polish
Semester of study	3	ECTS credits	2.0
Learning profile	general academic profile	Assessment form	exam

Conducting unit	Department of Biomedical Engineering -> Faculty of Electronics, Telecommunications and Informatics

Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Magdalena Mazur-Milecka
	Teachers	dr inż. Magdalena Mazur-Milecka

Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	0.0	15.0	0.0	30
	E-learning hours included: 0.0						

Learning activity and number of study hours	Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
	Number of study hours	30	2.0	18.0	50

Subject objectives	The aim of the course is to familiarize students with algorithms for explaining decisions of methods and networks of artificial intelligence.

Learning outcomes	Course outcome	Subject outcome	Method of verification
	[K7_W08] Knows and understands, to an increased extent, the fundamental dilemmas of modern civilisation, the main development trends of scientific disciplines relevant to the field of education.	The effect of the learning process is the student's gaining knowledge of modern methods of explaining decisions and trends in their development.	[SW1] Assessment of factual knowledge
	[K7_U01] can apply mathematical knowledge to formulate and solve complex and non-typical problems related to the field of study by:n-appropriate selection of source information and its critical analysis, synthesis, creative interpretation and presentation,n-application of appropriate methods and toolsn	The effect of the learning process is the student's acquisition of knowledge in the field of knowledge and the ability to apply measures for assessing the quality of neural networks and their reliability, as well as assessing the impact of features or parameters on the result.	[SU4] Assessment of ability to use methods and tools
	[K7_W04] Knows and understands, to an advanced extent, the principles, methods and techniques of programming and the principles of computer software development or programming devices or controllers using microprocessors or programmable elements or systems specific to the field of study, and organisation of systems using computers or such devices	The effect of the learning process is the student's acquisition of knowledge in the field of programming methods and techniques used in solving the problems of explaining decisions related to machine learning, including methods of visualization of weights, network parameters and the effect of features on the results.	[SW1] Assessment of factual knowledge

Subject contents	1. Introduction, introduction, basic concepts
	2. Clarification of decisions in artificial neural networks
	3.Quality assessment measures (IoU, mAI)
	4. Visualization of weights and parameters in models
	5. Assessment and visualization of the influence of features on the result in convolutional networks - introduction
	6. Assessment and visualization of the influence of features on the result in convolutional networks - CAM class methods
	7. Evaluation and visualization of Heatmaps, Layer-wise Relevance Propagation (LRP) methods
	8. Assessment and visualization of new methods of explaining decisions in artificial intelligence
	9. Assessment of the credibility of decision models and the quality of responses in artificial intelligence

Prerequisites and co-requisites	Basic knowledge of neural networks

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
lectures	50.0%	40.0%
lab	50.0%	60.0%

Recommended reading	Basic literature	1. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning Editors: **Samek**, W., **Montavon**, G., **Vedaldi**, A., **Hansen**, L.K., **Müller**, K., Springer 2019
		2. Hands-On Explainable AI (XAI) with Python, D. Rothman, Packt 2020
	Supplementary literature	Interpretable Machine Learning *A Guide for Making Black Box Models Explainable.*Christoph Molnar, 2021
	eResources addresses	Adresy na platformie eNauczanie:

Example issues/ example questions/ tasks being completed	
Work placement	Not applicable