



Subject card

Subject name and code	Machine Learning in Sound and Image Engineering, PG_00069758						
Field of study	Informatics, Electronics and Telecommunications, Biomedical Engineering, Biomedical Engineering, Biomedical Engineering, Space and Satellite Technologies, Automatic Control, Cybernetics and Robotics						
Date of commencement of studies	February 2025	Academic year of realisation of subject	2025/2026				
Education level	second-cycle studies	Subject group					
Mode of study	Full-time studies	Mode of delivery	at the university				
Year of study	1	Language of instruction	Polish				
Semester of study	2	ECTS credits	2.0				
Learning profile	general academic profile	Assessment form	assessment				
Conducting unit	Department of Multimedia Systems -> Faculty of Electronics Telecommunications and Informatics -> Faculties of Gdańsk University of Technology						
Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Adam Kurowski					
	Teachers	dr inż. Adam Kurowski dr hab. inż. Piotr Szczuko mgr inż. Szymon Zaporowski dr inż. Maciej Szczodrak					
Lesson types	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	15.0	0.0	0.0	30
	E-learning hours included: 0.0						
	eNauczanie source address: https://enauczanie.pg.edu.pl/2025/course/view.php?id=1876						
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	0.0	0.0	30		
Subject objectives	The aim of the course is to familiarize students with contemporary techniques of multimedia signal processing (using sound and image recording). These issues include knowledge of acoustic signal acquisition, physiology of the speech organ, speech signal analysis and synthesis, analysis and segmentation of visual signals, and illustrating the possibilities of machine learning applications, in particular the ability to adaptively learn and interpret possible variants of objects or patterns, e.g., scale, orientation, and perspective in the field of multimedia technologies. The course also presents tools in the form of scripts and code libraries for the Python language.						

Learning outcomes	Course outcome	Subject outcome	Method of verification
	[K7_W02] knows and understands, to an increased extent, selected laws of physics and physical phenomena, as well as methods and theories explaining the complex relationships between them, constituting advanced general knowledge in the field of technical sciences related to the field of study	The student is able to explain the phenomena of speech emission and perception, select the appropriate tools, and correctly perform signal recording and preprocessing for the purposes of machine learning applications.	[SW1] Assessment of factual knowledge
	[K7_W101] is able to make an in-depth identification of key objects and phenomena related to the field of study, as well as theories that describe them and applicable analytical and design methods	The student knows the methods of creating and organizing services in the field of machine learning and artificial intelligence and is able to model objects in an image and understand the concepts of segmentation and region of interest detection. The student knows the concepts of multi-task learning, including designing models with branches that perform different tasks simultaneously.	[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: - appropriate selection of source information and its critical analysis, synthesis, creative interpretation and presentation, - application of appropriate methods and tools	The student is able to select the appropriate tools for correct preprocessing of text and speech data, define the main problem and objective, and apply input data processing methods in order to achieve the intended goal.	[SU1] Assessment of task fulfilment [SU3] Assessment of ability to use knowledge gained from the subject
	[K7_U12] is able, to an increased extent, to analyze the operation of components and systems related to the field of study, as well as to measure their parameters and study their technical characteristics, and to plan and carry out experiments related to the field of study, including computer simulations, interpret the obtained results and draw conclusions	The student is able to critically analyze the machine learning algorithm training results and the application of machine learning methods dedicated to text processing and speech signal processing. The student can formulate conclusions from the completed work, from observations of the training process, and from the inference of selected deep models.	[SU1] Assessment of task fulfilment [SU2] Assessment of ability to analyse information [SU3] Assessment of ability to use knowledge gained from the subject [SU5] Assessment of ability to present the results of task
Subject contents	<p>Course content – lecture</p> <ul style="list-style-type: none"> • Digital signals, representations, parameterizations • Fundamentals of speech acoustics, signal acquisition, phonic devices, digital path, practical problems of signal recording, data set construction • Fundamentals of operation and applications of automatic speech recognition systems • Principle of operation and application of voice cloning solutions • Application of deep learning-based speech synthesizers for data augmentation • Speaker recognition using deep learning • Application of distance metric learning algorithms for visual and audio data • Object detection and semantic segmentation • Multitask learning in computer vision <p>Course content – laboratory</p> <ul style="list-style-type: none"> • Fundamentals of speech acoustics, signal acquisition, phonic devices, digital audio processing pipeline, practical problems of signal recording, audio data set construction • Use of deep learning-based speech synthesizers for speech synthesis • Overview of distance metric learning algorithms for visual data based on the analysis and comparison of contrastive loss and triplet loss algorithms • Multitask learning based on algorithms used in computer vision 		
Prerequisites and co-requisites	Students should have basic knowledge of data preparation in machine learning, know and distinguish between artificial intelligence methods, and have practical experience in using tools that enable neural network learning.		
Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	Trial exam	50.0%	50.0%
	Laboratory reports	50.0%	50.0%

Recommended reading	Basic literature	<ul style="list-style-type: none"> • Course materials posted on the eNauzanie website • Goodfellow Ian, Bengio Yoshua, Courville Aaron. Deep learning, Learning systems. PWN 2018. • François Chollet, Matthew Watson - Deep Learning with Python, Third Edition, Manning Publications, 2025 • Sebastian Raschka, Vahid Mirjalili - Python. Machine learning and deep learning. Scikit-learn and TensorFlow 2 libraries (Third Edition, 2024) • M. Cord, P. Cunningham, Machine Learning Techniques for Multimedia. Case Studies on Organization and Retrieval. Springer (Part of the Cognitive Technologies book series (COGTECH)) • K. Pardeep, Singh, Amit Kumar (Eds.) Machine Learning for Intelligent Multimedia Analytics Techniques and Applications. Springer ISBN 978-981-15-9492-220 • Dey, Nilanjan & Ashour, Amira S. & Nhu, Nguyen. (2016). Deep Learning for Multimedia Content Analysis. 10.1201/9781315399744-15.
	Supplementary literature	<ul style="list-style-type: none"> • Lecture materials for Sound and Image Engineering courses - https://sound.eti.pg.gda.pl/student/materialy.html • Wahiba Ben Abdessalem Karaa, Nilanjan Dey, Mining Multimedia Documents. CRC Press/Taylor & Francis Group, 2017, ISBN 11380317 • Lane Hobson, Cole Howard, Hannes Hapke, Natural Language Processing in Action, PWN 2021 • Le, L., Zheng, Y., Carneiro, G., Yang, L. (Eds.) Deep Learning and Convolutional Neural Networks for Medical Image Computing Precision Medicine, High Performance and Large-Scale Datasets. Springer • H. Purwins, B. Li, T. Virtanen, J. Schlüter, S. -Y. Chang and T. Sainath, Deep Learning for Audio Signal Processing, in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 2, pp. 206-219, May 2019, doi: 10.1109/JSTSP.2019.2908700
	eResources addresses	
Example issues/ example questions/ tasks being completed		
Practical activities within the subject	Not applicable	

Document generated electronically. Does not require a seal or signature.