Warsztaty

Łukasz Augustyniak

Temat: Large Language Models - from Demo to Production

Abstrakt: In this interactive workshop, we will delve into the fascinating world of large language models and explore their potential in revolutionizing various industries. I will guide participants through the process of transforming cutting-edge demos into scalable production solutions.

Participants will gain hands-on experience by working on practical exercises that demonstrate how to fine-tune these models for specific tasks. Additionally, we'll cover best practices for deploying these models at scale while maintaining efficiency and performance.

Throughout the workshop, attendees can expect engaging discussions about ethical considerations surrounding AI usage as well as insights into future developments within the field. By the end of this session, participants should have a solid understanding of how to harness the power of large language models effectively in order to drive innovation across various domains.

Arkadiusz Janz

Temat: Training Large Language Models with Reinforcement Learning from Human Feedback (RLHF)

Abstrakt: A comprehensive introduction to Generative Language Models and Reinforcement Learning from Human Feedback: a novel approach in training Large Language Models for downstream tasks. This workshop is designed to impart an in-depth understanding of fundamental concepts of Reinforcement Learning (states, actions, rewards, value functions, policies) and Generative Language Models. A theoretical comparison with Supervised Learning paradigm will be discussed, along with the advantages RLHF optimization brings to reducing biases, and overcoming sparse reward issues. Participants will engage in hands-on activities involving RLHF training with simplified models, hyperparameter tuning of RLHF models, and diving into existing RLHF programming frameworks.

Konrad Wojtasik

Temat: Introduction to modern information retrieval

Abstrakt: Information retrieval plays a crucial role in modern systems, finding applications across diverse domains and industries. Its relevance spans from web search and recommendation systems to product search and health and legal information retrieval. Information retrieval is not only essential for traditional search applications but also plays a vital role in retrieval-augmented Question Answering systems. Additionally, it serves as a valuable mechanism to prevent Large Language Models from generating incorrect or hallucinated information. Moreover, it ensures that their knowledge remains accurate and up-to-date. During this workshop, participants will have the opportunity to explore and understand current state-of-the-art models used in information retrieval. They will gain insights into the strengths and limitations of these models. Furthermore, the workshop will focus on setting up an information retrieval pipeline, allowing participants to gain hands-on experience in building and implementing such systems. Additionally, participants will learn how to effectively measure and evaluate the performance of their information retrieval pipelines.

Mateusz Gniewkowski

Temat: Model Agnostic Explanations Techniques

Abstrakt: Machine learning models can often be complex and difficult to understand, therefore it is important to be able to explain how these models work, as they are increasingly used in a wide range of industries and applications.

The workshop will start by discussing some basic ways to explain machine learning models, such as using feature importance measures, decision trees, and visualization tools. However, the focus will then shift to model-agnostic techniques, which can be applied to any type of machine learning model. The techniques that will be covered in the workshop include LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations). These libraries are designed to provide more transparent and understandable explanations for machine learning models, even when the models themselves are complex or difficult to interpret.

Piotr Bielak

Temat: Introduction to graph representation learning

Abstrakt: In recent years, representation learning has attracted much attention both in the research community and industrial applications. Learning representations for graphs is especially challenging due to the relational nature of such data, i.e., one must reflect both the rich attribute space and graph structure in the embedding vectors. During this workshop, I will show how to use the PyTorch-Geometric library to easily build graph representations and solve a variety of applications. We will explore node, edge and graph-level representations through the prism of their associated downstream tasks and corresponding deep learning models (Graph Neural Networks).

Denis Janiak

Temat: Bayesowski Deep Learning i reprezentacje

Uncertainty estimation is a critical aspect of artificial intelligence systems, enabling them to quantify their confidence and provide reliable predictions. However, accurately assessing uncertainty becomes increasingly challenging when AI models encounter scenarios outside their training data distribution. This workshop, titled "Does Representation Know What It Doesn't Know?," aims to explore the concept of uncertainty estimation in AI systems and delve into the question of whether representations within these systems possess the ability to recognize their own limitations. During the workshop, we will investigate the various techniques and methodologies employed in uncertainty estimation, such as Bayesian approaches and deep learning-based techniques. We will analyze the strengths and limitations of these approaches and discuss their implications for real-world applications.

Furthermore, the workshop will delve into the concept of representation learning and its impact on uncertainty estimation. We will examine whether AI systems can effectively recognize when they are faced with novel or out-of-distribution inputs. Additionally, we will explore approaches to measure and improve representation awareness, enabling models to identify areas of uncertainty and seek further guidance or human intervention when necessary.

By the end of the workshop, attendees will gain a deeper understanding of the state-of-the-art techniques for uncertainty estimation and its importance in building robust AI systems. They will also gain insights into the fundamental question of whether representations within AI models possess the capability to identify areas of uncertainty and adapt accordingly.

Albert Sawczyn

Temat: Uczenie reprezentacji grafów wiedzy

Knowledge graphs have emerged as powerful tools for organizing and representing structured information in various domains, enabling efficient data integration, inference, and knowledge discovery. Knowledge graph representation learning aims to capture the rich semantic relationships and contextual information within knowledge graphs, facilitating effective knowledge inference and reasoning. This workshop aims to introduce the fundamental challenge of learning representations for knowledge graphs and highlight their significance in practical applications. Practical demonstrations will show how to easily learn representation using the PyKEEN library and how to apply it to a real-world NLP problem.

Jakub Binkowski

Temat: Generative models for graphs

Abstrakt: After many advancements in the realm of Graph Representation Learning, graph generation gained much attention due to its vast range of applications (e.g. drug design). Nonetheless, due to the nature of graph data, this task is very difficult and further breakthroughs still need to be discovered. Hence the workshop will provide a ground understanding of the selected methods and problems associated with graph generation. During the workshop, I will show the most important methods in theory and practice. I will show how to implement these methods leveraging Pytorch Geometric library. We will go through training and evaluation using common datasets.

Kamil Kanclerz

Temat: Subjective problems in NLP

Abstrakt: A unified gold standard commonly exploited in natural language processing (NLP) tasks requires high inter-annotator agreement. However, there are many subjective problems that should respect users' individual points of view. At the first glance, disagreement and non-regular annotations can be seen as noise that drags the performance of NLP task detection models down. As we know, the ability to think and perceive the environment differently is natural to humans as such. Therefore, it is crucial to include this observation while building predictive models in order to reflect the setup close to reality. As simple as this may seem, it is important to keep in mind that the key ideas behind NLP phenomenon detection, such as gold standard, agreement coefficients, or the evaluation itself need to be thoroughly analyzed and reconsidered especially for subjective NLP tasks like hate speech detection, prediction of emotional elicitation, sense of humor, sarcasm detection, or even sentiment analysis. Such NLP tasks come with each complexity of their own, especially within the aspect of subjectivity, therefore making them difficult to solve compared to non-subjective tasks.

During the workshop, the participants will be introduced to the novel deep neural architectures leveraging various user representations. Moreover, the user-centered data setups will be explained in comparison to their ground truth equivalents. Additionally, the personalized evaluation techniques will be presented as the methods providing further insight into model ability to understand differences between various user perspectives.

Mateusz Wójcik

Temat: Continual Learning - techniques and applications

Abstrakt: Recently, neural architectures have become effective and widely used in various domains. The parameter optimization process based on gradient descent works well when the data set is sufficiently large and fully available during the training process. But what if we don't have all the data available during training? What if the number of classes increase? As a result, we have to manually retrain the models from scratch ensuing a time-consuming process.

During this workshop you will learn about the Continual Learning and its applications. We will discuss the catastrophic forgetting and explore various techniques that trying to prevent it starting from simple neural networks up to modern LLMs. As a result, you will understand why we need Continual Learning and how to apply it to existing or new models.

Patryk Wielopolski

Temat: Conditional object generation using pre-trained models and plug-in networks

Abstrakt: Generative models have gained many Machine Learning practitioners' attention in the last years resulting in models such as StyleGAN for human face generation or PointFlow for the 3D point cloud generation. However, by default, we cannot control its sampling process, i.e., we cannot generate a sample with a specific set of attributes. The current approach is model retraining with additional inputs and different architecture, which requires time and computational resources. During this hands-on workshop we will go through a method which enables to generate objects with a given set of attributes without retraining the base model. For this purpose, we will utilize the normalizing flow models - Conditional Masked Autoregressive Flow and Conditional Real NVP, and plug-in networks resulting in the Flow Plugin Network.

Mateusz Nurek

Temat: Complex networks I - social network analysis

Computational network science is a field of artificial intelligence that analyses graphs in applied problems involving social, transportation, epidemiological or energy issues. This workshop will teach you fundamental tools and techniques for analysing this data type. Based on a case study - the history of communication in a particular company, we will solve the problem of optimizing the structure of its organization. We will detect natural teams from employees most intensively working together. We will also identify key personnel, i.e. employees whose loss can cause communication paralysis.

Michał Czuba

Temat: Complex networks II - spreading processes

Abstrakt: Two years ago, the world faced SARS-CoV-2 and the biggest pandemic in the century. Since last winter, with an incursion of Russian troops in Ukraine, all civilised countries have been subjected to misinformation. This year with an election in Poland, a festival of campaign promises has started. The nature of these three examples is complex and hard to analyse. Nonetheless, one of the approaches leading to understanding and controlling such processes is a network simulation. During this workshop, you will learn an essential toolkit to model and analyse spreading phenomena in complex networks. You will understand how to simulate such processes as epidemics or opinion dynamics and how to identify key spreaders of fake news or the most fragile individuals to be vaccinated in the first place.

Michał Karol

Temat: Computer Vision for medical image processing

Abstrakt: Computer vision has emerged as a revolutionary technology in the medical field, bringing significant transformations in various aspects of healthcare. Its application in clinical practice has paved the way for improved diagnostics, more accurate disease detection, and enhanced treatment planning. The objective of this workshop is to bring comprehensive understanding of the impact of computer vision in clinical practice. Participants will gain insights into how this technology is reshaping healthcare and improving patient outcomes. By exploring the latest advancements in certified medical systems, attendees will learn about the integration of computer vision into existing medical frameworks and protocols. Moreover, the workshop will delve into current research areas within computer vision in medicine. Participants will be introduced to cutting-edge studies and ongoing projects that aim to further enhance the capabilities of computer vision in the healthcare domain. In the second part of the workshop, there will be an interactive session focused on implementing classification and segmentation networks using the JAX framework and the Flax library.

Piotr Kawa

Temat: Generating audio DeepFakes and how to detect them

Abstrakt: Recent advances in audio processing and speech synthesis allow the creation of realistic speech - DeepFakes. Despite a number of advantages, this technology poses a threat through its potential applications including disinformation spreading. Participants in this workshop will learn about the latest technologies for speech generation and then how to defend themselves against the malicious use of this technology using state-of-the-art DeepFake detection techniques.

Wojciech Wodo

Temat: Ekosystem tożsamości cyfrowej w Polsce i Europie (nowy e-dowód, eIDAS, SSI, EUDI Wallet)

Abstrakt: Podczas warsztatu naświetlę obecny stan ekosystemu tożsamości cyfrowej w Polsce i Europie oraz rozwinę jego przyszłość w związku z najnowszymi zmianami legislacyjnymi (eIDAS2), która zrewolucjonizuje rynek tożsamości cyfrowej poprzez wprowadzenie European Union Digital Identity Wallet (EUDI wallet), umożliwiając więcej funkcjonalności i dając kontrolę nad atrybutami i danymi użytkownikowi (posiadaczowi danych).

Ponadto skupię się na obecnych już na rynku środkach identyfikacji elektronicznej, takich jak nowy dowód osobisty z warstwą elektroniczną (e-dowód) oraz portfel tożsamości – mObywatel jak również usługi podpisów cyfrowych (podpis zaufany czy podpis osobisty).

Michał Walkowski

Temat: Wprowadzenie do analizy statycznej aplikacji mobilnych

Abstrakt: Monitorowanie systemu plików i analiza aktywności sieciowej mogą nam wiele powiedzieć o aplikacji. Natomiast są przypadki, w których musimy dokładniej ocenić funkcjonalność analizowanej przez nas aplikacji w celu weryfikacji zachowanie lub zidentyfikowania niepożądanego działania. Analiza statyczna koncentruje się na ocenie samego pliku wykonywalnego bez jego wykonywania, zwykle poprzez inżynierię wsteczną.

Inżynieria wsteczna nie wymaga od Ciebie bycia programistą, ale wymaga pewnych umiejętności analitycznych, czasu na badania i analizę oraz dużo czasu na przeszukiwanie Internetu w celu poszukiwania odpowiedzi na nurtujące nas pytanie.

Robert Czechowski

Temat: Informatyka śledcza - wybrane zagadnienia i narzędzia kryminalistyki cyfrowej

Abstrakt: Informatyka śledcza pod kątem procesu analizy jest wyjątkowa pod każdym względem. Oprócz wiedzy teoretycznej i umiejętności praktycznych wymagana jest spora wyobraźnia i umiejętność nieszablonowego podejścia do każdej analizowanej sprawy. Głównym celem informatyki śledczej oprócz dostarczenia cyfrowych materiałów dowodowych umożliwiających potwierdzenie lub zaprzeczenie, iż dane zdarzenie miało miejsce, jest również przedstawienie scenariusza i toku postępowania w danej sprawie (najczęściej przygotowawczej lub postępowania sądowego). Głównymi celami informatyków śledczych jest ujawnienie i rekonstrukcja zdarzeń mających charakter kryminalny, prowadzących do zakłócenia innych legalnych działań cyfrowych lub coraz częściej – cyfrowych przestępstw.

W trakcie warsztatów zostaną przedstawione:

- informacje, kto może zostać informatykiem lub funkcjonariuszem śledczym, oraz kto może zostać biegłym sądowym w zakresie kryminalistyki cyfrowej – i jakie predyspozycje należy posiadać,
- procesy akwizycji cyfrowego materiału dowodowego oraz sposoby i techniki analizy (narzędzia open-source),
- praktyczne zadania-wyzwania do samodzielnej realizacji.

W trakcie zajęć uczestnicy będą mieli okazję zapoznać się pracą informatyka śledczego oraz samodzielnie (lub z małą pomocą) zrealizować powierzone im zadania w formie logicznych zagadek.

WYKŁADY

Konrad Kałużny

Temat: Od źródeł threat intelligence do dowodów cyfrowych - Nowe strategie wykrywania i polowania na cyberzagrożenia

Abstrakt: Podczas tej sesji omówię podstawowe zasady procesów wykrywania zagrożeń oraz threat huntingu. Podzielę się kilkoma cennymi wskazówkami dotyczącymi źródeł threat intelligence i sposobów ich wykorzystania. Przyjrzymy się również różnym źródłom danych oraz dowodom cyfrowym pozostawionym po emulacji zagrożeń.

Jako case study skoncentrujemy się na najnowszych technikach ataku wykorzystanych przez cyberprzestępców w roku 2023. Dokładniej omówimy techniki, które stały się popularnymi wektorami ataku. Będziemy analizować ich mechanizmy działania oraz strategie, które przestępcy zastosowali, aby uniknąć wykrycia.

Poszerzymy wiedzę na temat aktualnych zagrożeń oraz omówimy kilka praktycznych wskazówek dotyczących identyfikacji tych ataków. Celem jest zwiększenie świadomości o konieczności proaktywnego podejścia do bezpieczeństwa oraz dostarczenie narzędzi i strategii, które pomogą w codziennej pracy oraz w zmaganiach z cyberprzestępczością.

Bogusław Szczupak, Mateusz Mądry

Temat: Hackowanie na bazie SDR

Abstrakt: Podczas wystąpienia omówiona zostanie obsługa i specyfikacja radia zdefiniowanego programowo (SDR). Zaprezentowany zostanie sposób analizy komunikacji radiowej oraz przeprowadzenia ataków z użyciem SDR. Omówione zostaną aspekty prawne użytkowania SDR.