Welcome to AIDD 2024
2ⁿᵈ International Conference on AI, Data Mining and Data Science (AIDD 2024)
October 05 ~ 06, 2024, Virtual Conference

Call for Participation

We invite you to join us in **2ⁿᵈ International Conference on AI, Data Mining and Data Science (AIDD 2024)**

This conference will provide an excellent international forum for sharing knowledge and results in theory, methodology and applications of Artificial Intelligence, Data mining, and Data Science. The Conference looks for significant contributions to all major fields of the AI, Data mining, and Data Science in theoretical and practical aspects.

Highlights of AIDD 2024 include:

- 2ⁿᵈ International Conference on Blockchain, Cybersecurity and Internet of Things (BCYIoT 2024)
- 2ⁿᵈ International Conference on Education & Integrating Technology (EDIT 2024)
- 2ⁿᵈ International Conference on Semantic Technology (SEMTEC 2024)
- 2ⁿᵈ International Conference on NLP, AI & Information Retrieval (NLAII 2024)
- 2ⁿᵈ International Conference on Clinical and Medical Sciences (CAMS 2024)
- 7ᵗʰ International Conference on Mechanical Engineering & Applications (MEAP 2024)
- 8ᵗʰ International Conference on Electrical Engineering (ELE 2024)
- 2ⁿᵈ International Conference on Vision Computing (VISCOM 2024)
- 2ⁿᵈ International Conference on Computer Science and Information Technology Advances (CCSITA 2024)

Registration Participants

Non-Author / Co-Author/ Simple Participants (no paper)

100 USD (With proceedings)

Here's where you can reach us: **aidd@aidd2024.org**(or) **aaiddconf@gmail.com**

Accepted Papers

BLOCK MEDCARE: ADVANCING HEALTHCARE THROUGH BLOCKCHAIN INTEGRATION

Oliver Simonoski1, Dijana Capeska Bogatinoska2, 1Faculty of Communication Networks and Security, University of Information Science and Technology, St. Paul the Apostle, Ohrid, 6000, Macedonia, 2Faculty of Applied IT, Machine Intelligence and Robotics, University of Information Science and Technology, St. Paul the Apostle, Ohrid, 6000, Macedonia

ABSTRACT

In an era driven by information exchange, transparency holds crucial importance, particularly within the healthcare industry, where data integrity and confidentiality are paramount. This paper delves into the landscape of blockchain technology, elucidating its potential applications and existing solutions. Specifically, the study focuses on its integration into digital healthcare services, with a primary emphasis on Electronic Health Records (EHR). Leveraging blockchain-based implementations, patients gain the ability to securely store their medical data, facilitated through smart contracts capable of executing key functions such as Registration, Data Append, and Data Retrieve. The research addresses challenges in implementing blockchain in healthcare, proposing a solution using digital signatures and Role-Based Access Control for enhanced security. Additionally, insights from a survey and the development of a blockchain-based application underscore the potential impact of blockchain on patient-centric and secure healthcare services, promising substantial contributions to the healthcare system. By deploying Ethereum-based blockchain implementations, patients gain the ability to securely manage their medical data through smart contracts, revolutionizing the way healthcare records are stored and accessed. This multi-layered approach ensures data integrity and controlled access. The findings of this research underscore the impact blockchain technology can have on healthcare solutions, indicating a new era of patient-centric and secure healthcare services.

OPENSTAXQA: A MULTILINGUAL DATASET BASED ON OPEN-SOURCE COLLEGE TEXTBOOKS

Pranav Gupta, Cornell University, Ithaca, New York, USA

ABSTRACT

We present OpenStaxQA, an evaluation bench- mark specific to college-level educational applications based on 43 open-source college text- books in English, Spanish, and Polish, available under a permissive Creative Commons license. We finetune and evaluate large language models (LLMs) with approximately 7 billion parameters on this dataset using quantized low rank adapters (QLoRa). Additionally, we also perform a zero-shot evaluation on the AI2 reasoning challenge dev dataset in order to check if OpenStaxQA can lead to an improved performance on other tasks. We also discuss broader impacts relevant to datasets such as OpenStaxQA.

BLOCKCHAIN APPLICATIONS IN CYBER LIABILITY INSURAN

Shreyas Kumar, Leah Kocian, and Luis Loo, Texas A&M University, College Station

ABSTRACT

Blockchain technology is revolutionizing cyber liability insurance (CLI) by addressing key challenges in underwriting, risk assessment, and claims management. As cyber-attacks become more frequent and complex, the demand for effective CLI solutions has surged. Traditional insurance practices often fall short in this rapidly evolving landscape. Blockchain offers a decentralized, secure, and transparent approach, enhancing the accuracy of risk assessments and preventing fraudulent claims. By maintaining an immutable ledger of historical claims,

blockchain allows for better comparison of new claims against past data. Additionally, smart contracts within blockchain frameworks can automate claims processing, reducing administrative tasks and speeding up resolutions. Blockchain also enables decentralized, peer-to-peer insurance platforms, allowing multiple insurers to pool resources and share risks in a transparent, efficient manner. This study explores how blockchain can transform CLI, improving efficiency and security across the industry.

OPCODE GUARD: FORTIFYING VULNERABILITY DETECTION WITH MACHINE LEARNING

Karishma Yadav, Malaviya National Institute of Technology, Jaipur, India

ABSTRACT
This research paper explores various smart contract opcode analysis techniques to detect software vulnerabilities and we will discuss about a theoretical machine learning-based model idea for opcode analysis. Opcode, the fundamental unit of machine-level instructions, is a valuable source of information to assess the security of binary and assembly code of smart contracts. Smart contracts use blockchain technology's decentralized and tamper-resistant nature to ensure transparency, security and immutability of the contract's execution. The paper provides a survey of static and dynamic analysis methods that leverage smart contract opcode sequences to identify potential security flaws. Researchers use opcode patterns and frequencies in static analysis to detect known vulnerabilities. In contrast, dynamic analysis techniques monitor the execution flow of smart contract opcodes during runtime to identify anomalies indicative of zero-day vulnerabilities. Machine learning model approaches is more accurate in investigating opcodes, where opcode frequency vectors or sequences are used to train models for vulnerability classification. This paper sheds light on the advancements, challenges, and prospects of continuously employing opcode-based techniques to pursue robust software security.

ENHANCING SMART CONTRACT SECURITY THROUGH OBFUSCATION: VERIFICATION VIA CONTROL FLOW GRAPH ANALYSIS

Karishma Yadav, Malaviya National Institute of Technology, Jaipur, India

ABSTRACT
Ethereum smart contracts leverage blockchain tech- nology to facilitate the transfer of values directly between participants on a network, eliminating the need for a central authority. These contracts are deployed on decentralized applications that operate on top of the blockchain. By doing so, they provide individuals with the ability to create agreements in a transparent and secure environment, minimizing conflicts and promoting trust. It has been observed that there are bugs in the smart contract's codes as these are provided by various programmers across the globe. The attackers exploit these security loopholes and pose a significant threat to applications, which subsequently result in financial losses to users. Discovering vulnerability in each contract is an important but time-consuming task. Therefore, we require to provide a security layer to each smart-contract such that it will make the exploitation a bit difficult task for attackers. The security layer is added by applying source-code obfuscation on smart contracts. The code obfuscation in security is mainly used by attackers to hide their malicious intent. we, in this approach suggest this method for increasing the complexity of smart contracts so that these cannot be exploited easily. We evaluate the impact of adding security layer to smart contract. The evaluation was done with various static and dynamic tools that identify the vulnerability in smart contracts. We achieved promising results which show that Obfuscation technique enhances the security and complexity of codes which are stored on public blockchain.

"IJUST CAN'T REMEMBER": LACKING A CLEAR VIEW – BRAIN FOG IN LONG COVID EDUCATION

Katharina F. Gallant, 1Center for Development Research, Bonn University, Bonn, Germany

ABSTRACT

When the COVID-19 pandemic struck in early 2020, the world was in crisis. Those countries who were sufficiently well off to focus on the pandemic focused on tackling the immediate health crisis. Long-term effects were feared in the realm of economics and, to some degree, in education. Long-term health risks limiting peoples' participation in all realms of public life were, however, hardly a point of concern. Today we see that the pandemic continues silently, not only in the form of new infections with COVID-19 but as people suffering from the syndrome of Long COVID. This paper addresses the situation of one of the least visible groups of Long COVID patients, that of children and adolescents, and discusses their needs within the field of school education

UNCOVERING GREENWASHING: A STUDY ON CORPORATE SUSTAINABILITY REPORTS AND PUBLIC SENTIMENT

Charlott Jakob

ABSTRACT

Greenwashing is an increasingly widespread problem in society. It is difficult to detect due to deceptive strategies and knowledge disparity between companies and the public. Companies can either present genuine or misleading statements about their efforts in their sustainability reports (SRs). Negative public opinion might be an incentive for companies to obscure certain statements in an attempt to counteract negative headlines. In this study, we investigate the link between a company's public image and its SR. Analysing the top 60 companies in Germany, we explore the connection between linguistic text features in their SRs and the sentiment towards them in media. The paper reveals a lack of significant feature differences between positive and negative sentiment. A single significant difference was found, indicating that companies with negative public sentiment use negations more frequently than companies with predominantly positive public sentiment. This research shows that while detecting signs of greenwashing is possible, it remains a challenging task.

FACILITATING STOCK RECOMMENDATIONS THROUGH SENTIMENT ANALYSIS

Shlok Bhura, Tanish Bhilare, Rylan Nathan Lewis and Dr. Kavita Kelkar, Department of Computer Engineering, K.J. Somaiya College of Engineering, Mumbai, India

ABSTRACT

Sentiment analysis is a relatively new method of stock recommendation that assesses news articles, social media feeds, and other information sources to ascertain investor sentiment towards a particular stock using machine learning and natural language processing. The model suggests whether to buy, hold, or sell the stock based on sentiment analysis. By emphasising trends and patterns in investor sentiment, the objective is to give investors insightful information that can help their decision-making. Several methods, including Decision Trees, Random Forests, Logistic Regression, and Gradient Boosting, were implemented to find the most accurate sentiment

analysis model. With an accuracy score of 85.02% among all, the Random Forest model came out as the most appropriate.

SIMILAR DATA POINTS IDENTIFICATION WITH LLM: A HUMAN-IN-THE-LOOP STRATEGY USING SUMMARIZATION AND HIDDEN STATE INSIGHTS

Xianlong Zeng1, Yijing Gao2, Ang Liu3, and Fanghao Song1, 1Electrical Engineering and Computer Science, Ohio University, Athens, OH 45701, 2Data Science Institute, Brown University, Providence, RI 02912, 3Global Urban Studies Program, Rutgers University, New Brunswick, NJ 08901

ABSTRACT
This study introduces a simple yet effective method for identifying similar data points across non-free text domains, such as tabular and image data, using Large Language Models (LLMs). Our two-step approach involves data point summarization and hidden state extraction. Initially, data is condensed via summarization using an LLM, reducing complexity and highlighting essential information in sentences. Subsequently, the summarization sentences are fed through another LLM to extract hidden states, serving as compact, feature-rich representations. This approach leverages the advanced comprehension and generative capabilities of LLMs, offering a scalable and efficient strategy for similarity identification across diverse datasets. We demonstrate the effectiveness of our method in identifying similar data points on multiple datasets. Additionally, our approach enables non-technical domain experts, such as fraud investigators or marketing operators, to quickly identify similar data points tailored to specific scenarios, demonstrating its utility in practical applications. In general, our results open new avenues for leveraging LLMs in data analysis across various domains.

CASE STUDY - FULLY RECOVERED AFTER 3.5 YEARS OF LONG COVID

Helga Gielen, Post- Long COVID patient, Belgium

ABSTRACT
Long COVID was a rather neurological/immunological disorder in this Post-Long COVID patient (end of Feb. '20 - Oct. '23). Promising positive results, via a new 'soft trial approach' of the autonomic nervous system out of balance, have led to a cure for the Long COVID symptoms with, among other things, PEM. This after 3.5 years. Since then, until today, September 2024, there has been no relapse. She has been 100% active again since the summer of 2023 as she was before the first Covid infection.

DAPLSR: DATA AUGMENTATION PARTIAL LEAST SQUARES REGRESSION MODEL VIA MANIFOLD OPTIMIZATION

Haoran Chen, Jiapeng Liu, Jiafan Wang, and Wenjun Shi, School of Computer Science and Technology, Zhengzhou University of Light Industry, Zhengzhou 450000, China

ABSTRACT

Traditional Partial Least Squares Regression (PLSR) models frequently underperform when handling data characterized by uneven categories. To address the issue, this paper proposes a Data Augmentation Partial Least Squares Regression (DAPLSR) model via manifold optimization. The DAPLSR model introduces the Synthetic Minority Over-sampling Technique (SMOTE) to increase the number of samples and utilizes the Value Difference Metric (VDM) to select the nearest neighbor samples that closely resemble the original samples for generating synthetic samples. In solving the model, in order to obtain a more accurate numerical solution for PLSR, this paper proposes a manifold optimization method that uses the geometric properties of the constraint space to improve model degradation and optimization. Comprehensive experiments show that the proposed DAPLSR model achieves superior classification performance and outstanding evaluation metrics on various datasets, significantly outperforming existing methods.

KEYWORDS
Partial Least Squares Regression, Data augmentation, Manifold optimization.

WTCL-DEHAZE:RETHINKING REAL-WORLD IMAGE DEHAZING VIA WAVELET TRANSFORM AND CONTRASTIVE LEARNING

Divine Joseph Appiah, Donghai Guan, and Mingqiang Wei, School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China

ABSTRACT
Images captured in hazy outdoor conditions often suffer from color distortion, low contrast, and loss of detail, which impair high-level vision tasks. Single image dehazing is essential for applications such as autonomous driving and surveillance, with the aim of restoring image clarity. In this work, we propose WTCL-Dehaze an enhanced semi-supervised dehazing network that integrates Contrastive Loss and Discrete Wavelet Transform (DWT). We incorporate contrastive regularization to enhance feature representation by contrasting hazy and clear image pairs. Additionally, we utilize DWT for multi-scale feature extraction, effectively capturing high-frequency details and global structures. Our approach leverages both labeled and unlabeled data to mitigate the domain gap and improve generalization. The model is trained on a combination of synthetic and real-world datasets, ensuring robust performance across different scenarios. Extensive experiments demonstrate that our proposed algorithm outperforms state-of-the-art single image dehazing methods on both benchmark datasets and real-world images.

KEYWORDS
Image Dehazing, Semi-Supervised Learning, Contrastive Loss, Discrete Wavelet Transform, Deep Learning.

ACCESSIBILITY TESTING

Om Prakashl, SMS Lucknow, India

ABSTRACT
Accessibility initiative is for promoting the applications, websites and other online material's accessibility for those who suffer from some disabilities like visibility or hearing. Such disabilities restrict such people from engaging in many arts, community, cultural, and leisure events because the content of these events is inaccessible to them, i.e. they cannot hear it or see it. Fortunately, there are many firms who work to make such events, applications or activities accessible for handicapped people. Accessible Events Promotions Ltd is one such concern whose aim is to put all accessible events in one place in an accessible format and also promote access improvements..

KEYWORDS
Accessibility, SDLC, Accessible Events Promotions Ltd, hearing impaired.