

## **Project Topic 1: Distributed Optimization Under Adversarial Nodes**

**Subject: Networking and Security.**

**Abstract and Expectations:** In recent years, the topic of distributed optimization has become a canonical problem in the study of networked systems. In this setting, a group of agents equipped with individual objective functions are required to agree on a state that optimizes the sum of these functions. As in the classical consensus problem, the agents can only operate on local information obtained from their neighboring agents, described by a communication network.

The project investigates the vulnerabilities of consensus-based distributed optimization protocols to nodes that deviate from the prescribed update rule, such as due to failures or adversarial attacks. It begins by characterizing fundamental limitations on the performance of any distributed optimization algorithm in the presence of adversaries.

In this project, it is expected that the concepts of distributed optimization will be reviewed and then the necessary and sufficient conditions will be obtained to reach consensus in the presence of malicious nodes. If you wish, you can implement the code of the article or test and evaluate existing similar codes.

### **References:**

- [1] S. Sundaram and B. Ghahserifard, "Distributed optimization under adversarial nodes," IEEE Transactions on Automatic Control, vol. 64, no. 3, pp. 1063-1076, 2018.
- [2] E. Hegazi, H. Sjolund, and A. A. Abidi, "A filtering technique to lower LC oscillator phase noise," IEEE Journal of Solid-State Circuits, vol. 36, no. 12, pp. 1921-1930, 2001.
- [3] M. Rabbat and N. Robert, "Distributed optimization in sensor networks," In Proc. of the 3rd international Symposium on Information Processing in Sensor Networks, 2004, pp. 20-27.
- [4] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," IEEE Transactions on Automatic Control, vol. 57, no. 1, pp. 90-104, 2011.

## **Project Topic 2: Integration of Blockchain and Auction Models**

**Subject: Web Systems and Security.**

**Abstract and Expectations:** In recent years, blockchain has gained widespread attention as an emerging technology for decentralization, transparency, and immutability in advancing online activities over public networks. As an essential market process, auctions have been well studied and applied in many business fields due to their efficiency and contributions to fair trade.

The opportunities of applying blockchain in auctions or enhancing blockchain using auctions have attracted many research and innovation activities. In this project, it is expected to answer the following questions through a systematic literature survey:

- 1) What are the characteristics of existing blockchain technologies and auction models?
- 2) How can blockchain technologies and auction models enhance each other?
- 3) What blockchain-based auction applications have been published, and how can these applications be classified?
- 4) What auction-based solutions have been proposed for enhancing blockchains?
- 5) What open challenges can we identify in the integration between blockchain and auction models?

### **References:**

- [1] Z. Shi, Cees de Laat, P. Grosso and Z. Zhao, "Integration of blockchain and auction models: A survey, some applications, and challenges," *IEEE Communications Surveys & Tutorials*, 2022.
- [2] I. A. Omar, H. R. Haya R. Hasan, Raja Jayaraman, Khaled Salah and Mohammed Omar, "Implementing decentralized auctions using blockchain smart contracts," *Technological Forecasting and Social Change*, vol. 168, pp.: 120786, 2021.
- [3] Shu, Yun, "Blockchain for security of a cloud-based online auction system," PhD diss., Auckland University of Technology, 2018.

## **Project Topic 3: In-Context Unlearning**

**Subject: Web Systems and Security.**

**Abstract and Expectations:** Machine unlearning, the process of efficiently removing the influence of specific training data from a trained model, has become increasingly important due to privacy regulations like the Right to be Forgotten. This is especially pertinent for Large Language Models (LLMs) given the copyright concerns they raise. However, achieving precise unlearning for large models is computationally challenging. Recent research has introduced algorithms that approximate data removal without retraining the model but rely on access to model parameters, which may not always be feasible, especially in API-accessible LLMs.

In this project, it is expected that a review of unlearning methods will be done and the results of the In-Context Unlearning model will be compared with other methods.

### **References:**

- [1] M. Pawelczyk, N. Seth, and L. Himabindu, "In-context unlearning: Language models as few shot unlearners," arXiv preprint arXiv:2310.07579, 2023.
- [2] S. Neel, A. Roth and S. Sharifi-Malvajerdi, "Descent-to-delete: Gradient-based methods for machine unlearning," In Algorithmic Learning Theory, pp. 931-962, PMLR, 2021.
- [3] J. Z. Di, J. Douglas, J. Acharya, G. Kamath, and A. Sekhari, "Hidden poison: Machine unlearning enables camouflaged poisoning attacks," In *NeurIPS ML Safety Workshop*, 2022.

## **Project Topic 4: Efficient topic partitioning of Apache Kafka for high-reliability real-time data streaming applications**

**Subject: Web Systems, Database, and Networking.**

**Abstract and Expectations:** Apache Kafka is an event streaming platform known for its ability to reliably handle high-volume real-time data exchange. It follows a producer–consumer (pub-sub) pattern, where producers generate data and consumers consume it for various purposes. In this project, a novel approach to optimize two crucial configuration parameters in Apache Kafka is presented: the number of partitions and brokers.

The expectation of the project is to review a methodology for modeling the topic partitioning process in Apache Kafka, with a focus on data-intensive real-time flows within fog computing environments for ML/AI analytics. The project aims to formulate an optimization problem that determines the optimal number of partitions based on application requirements and constraints.

### **References:**

- [1] T. P. Raptis, C. Cicconetti, and A. Passarella, "Efficient topic partitioning of Apache Kafka for high-reliability real-time data streaming applications," *Future Generation Computer Systems*, vol. 154, pp. 173-188, 2024.
- [2] K. Fizza, N. Auluck and A. Azim, "Improving the schedulability of real-time tasks using fog computing," *IEEE Transactions on Services Computing*, vol. 15, no. 1, pp. 372-385, 2019.
- [3] T. P. Raptis and A. Passarella, "On efficiently partitioning a topic in apache kafka," In Proc. 2022 *International Conference on Computer, Information and Telecommunication Systems (CITS)*, 2022, pp. 1-8. IEEE.

## **Project Topic 5: Diffusion Models for Time Series Applications**

**Subject: Web Systems.**

**Abstract and Expectations:** This project aims to provide a comprehensive overview of diffusion models, a family of generative models rooted in deep learning, which have gained prominence in cutting-edge machine learning research. Known for their ability to generate samples resembling observed data, diffusion models find extensive application in image, video, and text synthesis. In recent years, the concept of diffusion has been extended to time series applications, leading to the development of numerous powerful models.

In this project, first, the concepts of diffusion models should be reviewed, and then different types of forecasting, imputation, and generation methods should be compared.

### **References:**

- [1] L. Lin, L. Zhengkun, L. Ruikun, L. Xuliang, and Junbin Gao, "Diffusion models for time-series applications: a survey," *Frontiers of Information Technology & Electronic Engineering*, pp.1-23, 2023.
- [2] T. Yan, Z. Hongwei, Z. Tong, Z. Yufeng, and X. Yuanqing, "ScoreGrad: Multivariate probabilistic time series forecasting with continuous energy-based generative models," *arXiv preprint arXiv:2106.10121*, 2021.
- [3] K. Rasul, S. Abdul-Saboor, S. Ingmar, B. Urs, and V. Roland, "Multivariate probabilistic time series forecasting via conditioned normalizing flows," *arXiv preprint arXiv:2002.06103*, 2020.

## Project Topic 6: Privacy-Preserving Machine Learning on Blockchain

**Subject:** Web Systems, Security, and Database.

**Abstract and Expectations:** This project aims to investigate new ways of using blockchain technology as a secure and transparent record-keeping system. By integrating blockchain with machine learning, we aim to enhance data privacy by securely accessing training data sets and sharing raw data. Using cryptographic methods and decentralized agreement systems, we can collaboratively improve machine learning models on blockchain platforms, ensuring data security and reliability.

For this project, we aim to carefully review at least 8 recent articles about keeping privacy in machine learning when using blockchain. After thoroughly studying and analyzing these articles, you will write a detailed report summarizing the important findings, methods used, and new developments in this innovative field. Moreover, you can run a basic simulator to demonstrate the complex ideas explained in the literature.

### References

- [1] H. Kim, S. U. Kim, J.-W. Hwang and C. Seo, "Efficient Privacy- Preserving Machine Learning for Blockchain Network", *IEEE Access*, vol. 7, pp. 136481-136495, Jan. 2019.
- [2] W. Moulahi, I. Jdey, T. Moulahi, M. Alawida, and A. Alabdulatif, "A blockchain-based federated learning mechanism for privacy preservation of healthcare IoT data," *Comput. Biol. Med.*, vol. 167, p. 107630, 2023, DOI: <https://doi.org/10.1016/j.compbimed.2023.107630>.
- [3] Priyan Malarvizhi Kumar et al., "Blockchain-enabled Privacy Preserving of IoT Data for Sustainable Smart Cities using Machine Learning", in *Proc. of 14th International Conference on COMmunication Systems (COMSNETS)*, pp. 1-6, 2022.
- [4] V. Drungilas, E. Vaičiukynas and L. Čeponienė, "Towards Collaborative Privacy-preserving Machine Learning on Private Blockchain," *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, Aveiro, Portugal, 2023, pp. 1-4, DOI: 10.23919/CISTI58278.2023.10211561.

## Project Topic 7: Netflix Prize Challenge Remake with Modern Collaborative Filtering Techniques

**Subject:** Web Systems, Database.

**Abstract and Expectations:** The Netflix Prize Challenge, a famous competition that aimed to improve the recommendation algorithm for Netflix, has inspired numerous advancements in collaborative filtering algorithms.

In this project, you will revisit the Netflix Prize Challenge concept and apply modern collaborative filtering techniques to enhance recommendation systems. The main aim of this project is to carefully read at least 8 recent academic articles about collaborative filtering algorithms. These articles will cover different methods like matrix factorization, deep learning, and hybrid models. You will also implement and compare different collaborative filtering methods using a dataset similar to the one used in the original Netflix Prize Challenge. We also expect to see an evaluation of the performance of each model in terms of accuracy, scalability, and computational efficiency. Finally, you will provide a detailed report on your findings, discussing the strengths and limitations of each approach and potential extensions for further improvement.

### References

- [1] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," in *Computer*, vol. 42, no. 8, pp. 30-37, 2009.
- [2] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. Chua, "Neural collaborative filtering," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 173-182, 2017.
- [3] H. Wang, N. Wang, D.Y. Yeung, and D. Yeung, "Collaborative deep learning for recommender systems," in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1235-1244, 2015.
- [4] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," in *ACM Computing Surveys (CSUR)*, vol. 52, no. 1, pp. 1-38, 2019.

## Project Topic 8: Personalized Book Recommendation System with NLP

**Subject: Web Systems and Database.**

**Abstract and Expectations:** Personalized book recommendation systems utilize Natural Language Processing (NLP) techniques to analyze user preferences, reading habits, and book content to provide tailored book suggestions to individuals. The project focuses on combining Natural Language Processing techniques with book data to create a personalized book recommendation system. NLP helps in extracting insights from book summaries and reviews, enabling the system to offer tailored suggestions to users based on their preferences. Your initial task is to delve into existing research algorithms related to book recommendation systems utilizing NLP. analyze and review at least 8 papers in this domain, synthesizing the findings and key insights into a comprehensive report for further study and reference in your project development. Subsequently, the project will involve the test of a novel algorithm within the realm of Big Data for testing the efficiency and scalability of the personalized book recommendation system.

### References

- [1] D. Sarma, T. Mittra, and M.S. Hossain, "Personalized Book Recommendation System using Machine Learning Algorithm," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 1, pp. 1, 2021. DOI: 10.14569/IJACSA.2021.0120126.
- [2] S. Kanetkar, A. Nayak, S. Swamy and G. Bhatia, "Web-based personalized hybrid book recommendation system," in *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, Unnao, India, 2014, pp. 1-5. DOI: 10.1109/ICAETR.2014.7012952.
- [3] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep Learning based Recommender System: A Survey and New Perspectives," *ACM Comput. Surv.*, vol. 1, no. 1, Article 1, July 2018, pp. 35.
- [4] W. Guo, H. Gao, J. Shi, B. Long, L. Zhang, B.-C. Chen, and D. Agarwal, "Deep Natural Language Processing for Search and Recommender Systems," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*, Association for Computing Machinery, New York, NY, USA, 2019, pp. 3199-3200. DOI: 10.1145/3292500.3332290.



## Project Topic 9: Opinion Mining on Product Reviews Using Big Data Processing

**Subject:** Database, Web Systems, and Networking.

**Abstract and Expectations:** Opinion mining, also known as sentiment analysis, delves into the sentiments expressed towards any entity. Opinion mining is about understanding what a text says and how it feels about the topic. As big data becomes more important, there's a growing need to use sentiment analysis in different areas, especially in product reviews. The current sentiment analysis models, even those used on social media like Twitter, and traditional models often depend a lot on old data, which slows down decision-making.

In this project, you're tasked with exploring diverse opinion mining methodologies. The primary goal is to go over a minimum of 8 recent articles and collect extensive product review datasets for analysis, focusing on sentiments associated with different products. We also want you to try out one of the opinion mining methods and see how accurate it is.

### References

- [1] A.P. Rodrigues and N.N. Chiplunkar, "Mining online product reviews and extracting product features using unsupervised method," in *2016 IEEE Annual India Conference (INDICON)*, Bangalore, India, 2016, pp. 1-6. DOI: 10.1109/INDICON.2016.7839054.
- [2] P. Baxi, "Sentiment Analysis Based on Social Networks Using Support Expectation-Maximization for E-Commerce Applications," in *2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Ballar, India, 2023, pp. 1-6. DOI: 10.1109/ICDCECE57866.2023.10150703.
- [3] M. Çataltaş, S. Doğramacı, S. Yumuşak, and K. Öztoprak, "Extraction of Product Defects and Opinions from Customer Reviews by Using Text Clustering and Sentiment Analysis," *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA, 2020, pp. 4529-4534. DOI: 10.1109/BigData50022.2020.9377851.
- [4] H.-C. Soong, N.B.A. Jalil, R.K. Ayyasamy, and R. Akbar, "The Essential of Sentiment Analysis and Opinion Mining in Social Media: Introduction and Survey of the Recent Approaches and Techniques," in *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Malaysia, 2019, pp. 272-277. DOI: 10.1109/ISCAIE.2019.8743799.

## **Project Topic 10 : Predictive Maintenance System Leveraging Sentiment Analysis on Industrial IoT Big Data**

**Subject: Database, Web Systems, and Networking.**

**Abstract and Expectations:** The development of a predictive maintenance system that integrates sentiment analysis on vast volumes of big data gathered from Industrial Internet of Things (IIoT) sensors holds significant promise for enhancing machinery maintenance processes in industrial settings. Prompted by the ongoing need for proactive maintenance solutions, this research aims to delve into at least 8 recently published papers on the subject.

The project will involve a comprehensive review and analysis of those papers, culminating in a detailed report assessing the state-of-the-art techniques in leveraging sentiment analysis for predictive maintenance in IIoT environments. Subsequently, one selected algorithm will be tested on the collected data to analyze its performance, and the findings will be thoroughly documented in the final report.

### **References**

- [1] S. Latif, M. Driss, W. Boulila, Z. Huma, S.S. Jamal, Z. Idrees, J. Ahmad, "Deep Learning for the Industrial Internet of Things (IIoT): A Comprehensive Survey of Techniques, Implementation Frameworks, Potential Applications, and Future Directions," *Sensors*, vol. 21, no. 22, pp. 7518, 2021. DOI: 10.3390/s21227518.
- [2] Sherien Elkateb, Ahmed Métwalli, Abdelrahman Shendy, Ahmed E.B. Abu-Elanien, "Machine learning and IoT – Based predictive maintenance approach for industrial applications," *Alexandria Engineering Journal*, vol. 88, no. 4, pp. 298-309, 2024, ISSN 1110-0168.
- [3] A.T. Keleko, B. Kamsu-Foguem, R.H. Ngouna, et al., "Artificial intelligence and real-time predictive maintenance in industry 4.0: a bibliometric analysis," *AI Ethics*, vol. 2, pp. 553-577, 2022.
- [4] Iantovics, L.B., Gligor, A., Montequín, V., Balogh, Z., Budinska, I., Gatial, E., Carrino, S., Ghorbel, H., Dreyer, J., "SOON: Social Network of Machines Solution for Predictive Maintenance of Electrical Drive in Industry 4.0," *Acta Marisiensis. Seria Technologica*, vol. 19, pp. 12-19, 2022. DOI: 10.2478/amset-2022-0012.

## Project Topic 11: Chat with documents

**Subject: Web Systems and Database.**

**Abstract and Expectations:** The advent of large language models (LLMs) has revolutionized the field of natural language processing. These LLMs, while not applications in themselves, are pre-trained statistical models that must be paired with an application (and, in some cases, specific data sources) in order to meet their purpose. For example, Chat-GPT is not an LLM, rather it is a chatbot application that utilizes GPT-3.5 or GPT-4 LLMs, depending on the selected version. LangChain is essentially a library that chains together different components to build applications on top of Language models.

In this project, firstly you're expected to review the key properties and methods used in LLMs e.g. few-shot learning and prompt engineering. If you wish, you can develop an interactive chatbot. This chatbot leverages public Persian LLMs and LangChain. All scripts should be written in Python and you can use Streamlit to create an interface.

## References

- [1] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, William Fedus "Emergent abilities of Large Language Models" Transactions on Machine Learning Research (TMLR), 2022
- [2] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, Ji-Rong Wen "A Survey of Large Language Models", preprint arxiv, 2023
- [3] Jean Kaddour, Joshua Harris, Maximilian Mozes, Herbie Bradley, Roberta Raileanu, Robert McHardy "Challenges and applications of Large Language Models", preprint arxiv, 2023
- [4] <https://learn.deeplearning.ai>, [LangChain for LLM Application Development]