ARTIFICIAL INTELLIGENCE CONFERENCE

BOOK OF ABSTRACTS

BELGRADE, OCTOBER 9-10, 2025

PROGRAM COMMITTEE

Co-chairs:
Vladan Devedžić, SASA, Belgrade and Faculty of Organizational Sciences, University of Belgrade
Miodrag Mihaljević, SASA, Belgrade and Mathematical Institute, SASA, Belgrade
Dušan Teodorović, SASA. Belgrade

Members:
Aleksandar Kostić, SASA, Belgrade
Angelo Sifaleras, University of Macedonia, Greece
Boško Nikolić, School of Electrical Engineering, University of Belgrade
Basant Agarwal, Central University of Rajasthan, Ajmer, India
Danijela Milošević, Faculty of Technical Sciences in Čačak, University of Kragujevac
Goce Trajčevski, Iowa State University, Ames, USA
Igor Đurović, Montenegro Academy of Sciences and Arts, Podgorica, Montenegro
Jelena Jovanović, Faculty of Organizational Sciences, University of Belgrade
Jesus Medina, University of Cadiz, Cadiz, Spain
Jianxun Cui, Harbin Institute of Technology, Harbin, China
Leonid Stoimenov, Faculty of Electronic Engineering, Univeristy of Niš
Marija Mitrović Dankulov, The Institute of Physics Belgrade
Mart Laanpere, Tallinn University, Tallinn, Estonia
Michael Benedikt, University of Oxford, Oxford, UK
Mirjana Ivanović, Faculty of Sciences, University of Novi Sad
Mirko Dinulović, Faculty of Mechanical Engineering, University of Belgrade
Predrag Janičić, Faculty of Mathematics, Univeristy of Belgrade
Ranka Stanković, Faculty of Mining and Geology, University of Belgrade
Slobodan Vukosavić, Faculty of Electrical Engineering, University of Belgrade and SASA, Belgrade
Sonja Dimitrijević, "Mihajlo Pupin" Institite, Belgrade
Vladimir Srdanović, Mathematical Institute SASA, Belgrade
Weiqin Chen, Oslo Metropolitan University, Oslo, Norway
Zoran Ognjanović, Mathematical Institute SASA, Belgrade

ORGANIZING COMMITTEE

Co-chairs:
Andreja Tepavčević, Mathematical Institute SASA, Belgrade and Faculty of Sciences, University of Novi Sad
Tatjana Davidović, Mathematical Institute SASA, Belgrade

Members:
Ana Anokić, Academy of Technical and Art Applied Studies, Belgrade
Anastazia Žunić, Mathematical Institute SASA, Belgrade
Anđelka Zečević, Mathematical Institute SASA, Belgrade
Biljana Stojanović, Mathematical Institute SASA, Belgrade
Božo Vuković, Mathematical Institute SASA, Belgrade
Dejan Vukelić, Mathematical Institute SASA, Belgrade
Dragan Jočić, Mathematical Institute SASA, Belgrade
Nataša Stević, SASA, Belgrade
Petar Vranić, Mathematical Institute SASA, Belgrade
Slobodan Jelić, Faculty of Civil Engineering, Univeristy of Belgrade
Una Stanković, Mathematical Institute SASA, Belgrade
Vanja Stepanović, Faculty of Agriculture, University of Belgrade

ORGANIZERS
Department of Technical Sciences SASA, Belgrade
Mathematical Institute SASA, Belgrade
Academy Board for Artificial Intelligence, SASA, Belgrade

FOREWARD

Following the two successful Artificial Intelligence Conferences held in December 2023 and December 2024, we have decided to organize a new edition, this time as an international conference. The Program Committee has been expanded to include nine researchers from nine different countries, representing diverse regions of the world.

The Artificial Intelligence Conference was initiated by a large community of AI researchers in Serbia, who wished to establish, under the auspices of the Serbian Academy of Sciences and Arts (SASA), the country's most prestigious scientific institution, a meeting primarily focused on research in this field.

The main goal of the conference is to bring together researchers from all areas of artificial intelligence, both from Serbia and abroad, to present their studies and findings. Given that artificial intelligence today is not only highly popular but also remarkably broad, the range of topics covered by this conference is correspondingly extensive.

In addition to emphasizing its research character, the significant features of this conference are:

• focusing exclusively on the field of artificial intelligence

• encouraging discussions on fundamental aspects of artificial intelligence

• paying special attention to new topics in artificial intelligence

• connecting techniques and results from various subfields of artificial intelligence

• bringing together a large number of researchers in the field of artificial intelligence from

Serbia and abroad, both those with experience and young researchers.

A total of 72 abstracts were submitted to the conference from all major universities and research institutions in Serbia and some universities abroad, covering a wide range of topics in the field of artificial intelligence. Of these, 68 submissions were selected for presentation. Each abstract was reviewed by an external reviewer or a member of the Program Committee, and several were accepted only after a second round of revision.

Artificial intelligence, with its profound and far-reaching influence on modern society, demands ongoing research and analytical reflection. This conference was conceived precisely with that goal in mind, to foster the exchange of research ideas and experiences in this broad and complex domain.

The organizers place particular emphasis on providing opportunities for young and promising researchers. Through presentations and discussions, young scholars will have the chance to showcase their work, receive valuable feedback from established experts, and form connections with potential future collaborators. We strongly believe that such intergenerational interaction is essential for the continuous advancement of artificial intelligence research in Serbia.

The organizers, representing three institutions, including the Department of Technical Sciences and Academy Board for Artificial Intelligence of the Serbian Academy of Sciences and Arts and the Mathematical Institute of the Serbian Academy of Sciences and Arts, are deeply convinced of the crucial role that artificial intelligence will play in shaping the future of many sectors in our country. We believe that research and innovation in this area will contribute to economic development, improvements in healthcare and education, and the creation of new perspectives in science, art, and other areas of life. With joint efforts, we can help shape the future of this vital field and promote the growth of knowledge, creativity, and innovation.

Sincerely,

Organizers of the Artificial Intelligence Conference,

Department of Technical Sciences SASA,

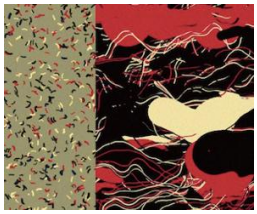
Mathematical Institute SASA, and

Academy Board for Artificial Intelligence SASA

Belgrade, October 2025

Legend:

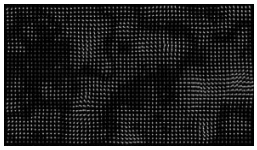
ARTIFICIAL INTELLIGENCE IN EDUCATION



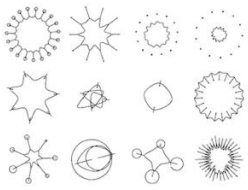
ARTIFICIAL INTELLIGENCE AND ETHICS, PRIVACY AND SOCIETY

ARTIFICIAL INTELLIGENCE IN SECURITY

COMPUTER VISION

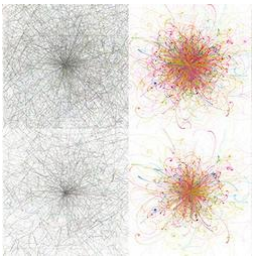


DECISION MAKING AND PROBLEM SOLVING

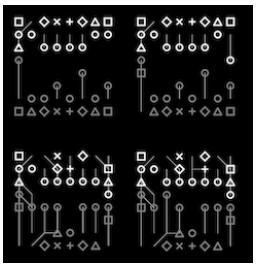


EXPLAINABLE ARTIFICIAL INTELLIGENCE

GENERAL AND GENERATIVE ARTIFICIAL INTELLIGENCE



KNOWLEDGE REPRESENTATION, UNCERTAIN KNOWLEDGE AND REASONING



MACHINE LEARNING

NATURAL LANGUAGE PROCESSING

KEYNOTE TALKS

AI in Health Care Strategies

Álvaro Rocha, ISEG, University of Lisbon, Portugal

Abstract: Artificial Intelligence (AI) is fundamentally reshaping healthcare, promising unprecedented advancements in medical diagnosis, treatment personalization, patient monitoring, and operational efficiencies. As healthcare systems globally face challenges such as aging populations, increasing chronic diseases, and resource limitations, AI emerges as a strategic tool to enhance care quality, reduce costs, and improve patient outcomes. This keynote explores the integration of AI into healthcare strategies, analyzing real-world applications, potential benefits, ethical considerations, and regulatory challenges. By examining case studies from clinical decision support systems, predictive analytics, telemedicine, and personalized medicine, we will highlight best practices and critical success factors for effective AI deployment. The session will also address practical considerations for healthcare leaders and policymakers, including data governance, patient privacy, and ethical frameworks necessary to build trust and ensure equitable healthcare outcomes.

Biography: Álvaro Rocha is World's Top 1% Scientist by Stanford University (USA) and Elsevier, World's Top 0.05% Scientist by ScholarGPS, and World's Top 1% Scientist by ResearchGate for the domains of Information Science, Information Systems and Business Informatics. He is a Professor of Information Systems at the University of Lisbon – ISEG and Honorary Professor at the Amity University, researcher at the ADVANCE (the ISEG Centre for Advanced Research in Management), and a collaborator researcher at the CINTESIS (Center for Research in Health Technologies and Information Systems). His main research interests are maturity models, management information systems, quality of information systems, intelligent information systems, cybersecurity, e-government, e-health, and information technology in education. He is also Vice-Chair of the IEEE Portugal Section Systems, Man, and Cybernetics Society Chapter, and Founder and Editor-in- Chief of two Scopus and SCIMago journals: JISEM - Journal of Information Systems Engineering & Management; and RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação / Iberian Journal of Information Systems and Technologies. Additionally, he is the Scientific Manager of the Information Systems Engineering & Management book series at Springer-Nature, the world's leading publisher of publications in Science, Technology and Health. Moreover, he has served as Vice- Chair of Experts for the European Commission's Horizon 2020 Program, and as an Expert at the COST - intergovernmental framework for European Cooperation in Science and Technology, at the European Commission's Horizon Europe Program, at the Government of Italy's Ministry of Universities and Research, at the Government of Latvia's Ministry of Finance, at the Government of Mexico's National Council of Science and Technology, at the Government of Polish's National Science Centre, at the Government of Cyprus's Research and Innovation Foundation, and at the Government of Slovak's Research Agency.

LLMs as GNNs (to understand how they generalise)

Petar Veličković, Senior Staff Research Scientist Google DeepMind, Affiliated Lecturer at the University of Cambridge, and Associate of Clare Hall, Cambridge

Abstract: Problems such as efficient (length) generalisation are highly relevant for building agentic systems, yet they have puzzled LLM researchers for several years. However, they can sometimes be obvious through the lens of graph representation learning. This is because LLMs are, by their design, vulnerable to generalisation issues from a variety of geometric perspectives. In this talk, we will leverage this insight, and attempt to scale the mountain of LLM generalisation. 🏔

Biography: Petar Veličković is a Senior Staff Research Scientist at Google DeepMind, Affiliated Lecturer at the University of Cambridge, and an Associate of Clare Hall, Cambridge. He holds a PhD in Computer Science from the University of Cambridge (Trinity College), obtained under the supervision of Pietro Liò. His research concerns aligning neural networks to (classical) computation, to assess and improve their out-of-distribution generalisation ability. Particularly, he focuses on neural algorithmic reasoning (featured in VentureBeat), graph representation learning and categorical and geometric deep learning (a topic he has co-written a proto-book about). For his contributions, he is recognised as an ELLIS Scholar in the Geometric Deep Learning Program. He is the first author of Graph Attention Networks - a popular convolutional layer for graphs-and Deep Graph Infomax-a popular self-supervised learning pipeline for graphs (featured in ZDNet). His research has been used in substantially improving travel-time predictions in Google Maps (featured in Endgadget, VentureBeat, CNET, The Verge and ZDNet), guiding intuition of mathematicians towards new top-tier theorems and conjectures (featured in Nature, Science, Quanta Magazine, New Scientist, The Independent, Sky News, The Sunday Times, la Repubblica and The Conversation), the first full AI system for tactical suggestions in association football (featured in Financial Times, The Athletic, The Economist, New Scientist, Wired, MIT Technology Review, The Verge and El País), and the first AI-assisted top-percentile result in competitive programming.

CONTRIBUTED TALKS

ARTIFICIAL INTELLIGENCE IN EDUCATION

From substitute to cognitive amplifier: The AIEB framework (AI as Educational Booster) for strategic teacher training

Matteo Ciastellardi[1], Giovanna Di Rosario [2], Paolo Ferri[3]

[1]Department of Design, Politecnico di Milano, Italy, matteo.ciastellardi@polimi.it
[2]Department of Human Sciences for Education Riccardo Massa, Università degli Studi di Milano Bicocca, Italy. giovanna.dirosario@unimib.it
[3]Department of Human Sciences for Education Riccardo Massa, Università degli Studi di Milano Bicocca, Italy, paolo.ferri@unimib.it

The emergence of generative AI represents a watershed moment in educational technology, comparable to the introduction of the internet in classrooms but potentially far more transformative in its cognitive implications. Since the public release of GPTs apps in November 2022, we have witnessed an unprecedented acceleration in AI adoption across educational contexts, from primary schools to universities. This rapid infiltration has occurred largely without systematic pedagogical frameworks, resulting in a paradoxical situation: while generative AI tools possess extraordinary potential to enhance human cognitive capabilities, they are predominantly employed as shortcuts or substitutes for critical thinking, creative processes, and deep learning. This misalignment between technological potential and pedagogical practice threatens to undermine the very foundations of education, transforming institutions of learning into spaces of cognitive outsourcing rather than intellectual empowerment.

Our research emerges from this critical context, proposing a paradigm shift in how educators conceptualize and integrate generative AI. Building on Vygotsky's (1978) zone of proximal development and Engelbart's (1962) framework of augmenting human intellect, we argue that generative AI should function as a cognitive amplifier—a tool that extends rather than replaces human capabilities. This perspective aligns with contemporary theories of distributed cognition (Hutchins, 1995) and the concept of "intelligence augmentation" rather than artificial intelligence as replacement. The distinction is crucial: while a substitute performs tasks instead of humans, a cognitive booster enhances human capacity to perform more complex, creative, and meaningful tasks.

This study presents findings from an innovative operational model developed through extensive field research with over 300 educators across five Italian regions. Our research capitalizes on the unique opportunity provided by Italy's PNRR investment in educational preparation, which enabled numerous schools to fund comprehensive teacher training programs. As one of the training groups, we had unprecedented access to diverse educational contexts, from urban technical institutes to rural primary schools, allowing us to develop and refine our AIEB (AI as Educational Booster) framework through iterative cycles of implementation and feedback.

The AIEB framework integrates three foundational pillars that transform generative AI from a passive tool into an active partner in the educational process. First, prompt literacy emerges as a new essential competency, extending beyond mere technical skills to encompass critical thinking about AI interactions. We conceptualize prompt design not as simple query formulation but as pedagogical architecture—each prompt becomes a scaffold that guides learners through increasingly complex cognitive territories. Second, the framework focuses on amplifying existing teacher competencies rather than replacing them. Through structured workshops, educators learned to leverage AI to enhance their subject expertise, pedagogical creativity, and assessment capabilities. Third, the model emphasizes concrete didactic and project-based applications, ensuring that AI integration translates into tangible improvements in lesson planning, student engagement, and learning outcomes.

A distinctive feature of our research is the development of the "Atlas of Knowledge for Generative AI in Educational Contexts"—a dynamic, self-updating repository that educators can use for continuous

professional development. This atlas includes practical consultation tools, self-assessment rubrics for prompt literacy, case studies of successful AI integration, and templates for AI-enhanced lesson planning. The atlas operates on principles of collective intelligence, with educators contributing their experiences and innovations, creating a living document that evolves with the technology and pedagogical practices.

Our preliminary findings reveal transformative shifts in educational practice. Teachers who completed the full AIEB training program demonstrated marked improvements in several key areas: a 52% increase in designing cognitively challenging tasks that promote higher-order thinking; a 78% shift from viewing AI as a time-saving device to recognizing it as a tool requiring sophisticated pedagogical integration; and a 41% improvement in identifying and supporting students with learning difficulties through AI-assisted diagnostic tools while maintaining essential human connection. Perhaps most significantly, we observed the emergence of "prompt progressions"—carefully sequenced AI interactions designed to gradually build student autonomy and critical thinking skills.

The implications of this research extend beyond immediate classroom applications. As we move forward, several critical areas demand continued investigation and development. The rapid evolution of AI capabilities necessitates ongoing updates to prompt literacy frameworks and integration strategies. We anticipate that future iterations of the AIEB framework will need to address emerging challenges such as multimodal AI integration, multi-agents in educational design, ethical considerations in AI-student interactions, and the development of AI-resistant assessment methods that ensure authentic learning. Our research opens pathways for longitudinal studies examining the long-term impact of AI as cognitive amplifier on student learning outcomes, teacher professional identity, and the evolution of educational institutions. The framework provides a foundation for international comparative studies and cross-cultural adaptations, recognizing that effective AI integration must be contextualized within specific educational traditions and values.

Keywords: Generative AI, Cognitive Amplification, Prompt Literacy, Teacher Professional Development, AIEB Framework

References:

[1] Engelbart, D. C. (1962). *Augmenting human intellect: A conceptual framework*. Stanford Research Institute.

[2] Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

[3] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

[4] Xu, H., Gan, W., Qi, Z., Wu, J., & Yu, P. S. (2024). *Large language models for education: A survey*. arXiv. https://doi.org/10.48550/arXiv.2405.13001

[5] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27.

AI-Powered Grading for Higher Education

Svetozar Ikovic[1], Jelena Graovac[2]

[1]Faculty of Mathematics, University of Belgrade,Serbia, mi241072@alas.matf.bg.ac.rs
[2]Faculty of Mathematics, University of Belgrade,Serbia, jelena.graovac@matf.bg.ac.rs

The rapid rise of Artificial Intelligence (AI) is reshaping many areas of society, and education is no exception. In particular, the growing use of Large Language Models (LLMs) is beginning to transform how teaching and assessment are approached. One area where this shift is especially promising is in the automated grading of student writing, particularly open-ended or essay-style responses [1][2][3]. Traditional grading of such responses is not only time-consuming but also prone to subjectivity and inconsistency among different graders. These challenges are especially pronounced in higher education, where instructors often face large volumes of student work. By leveraging the natural language understanding and generation capabilities of LLMs, educators can achieve faster, more consistent, and scalable assessments, while also improving the quality and timeliness of feedback provided to students. In addition to supporting instructors, this approach can also serve as a powerful learning aid for students. It offers an effective means of exam preparation through the creation of practice tests and self-assessments evaluated by AI, allowing learners to receive immediate, personalized feedback and better understand their progress.

In this work, we present a platform that automates the grading of textual student responses using OpenAI's LLMs. The system is implemented as a web application with a React frontend and a Django backend. Student submissions are processed in the backend, where they are combined with instructor-provided model answers or lecture materials and sent as structured prompts to the LLM. The output consists of both a grade and formative feedback, which are stored in the database and made accessible through the instructor-facing interface for review and adjustment. The modular architecture also supports extensions such as additional LLM providers, multilingual grading, and explainability features. The platform supports two complementary grading approaches: reference-based grading, in which student answers are evaluated against instructor-provided model responses; and generative grading, where the AI generates reference answers based on teaching materials such as lecture notes or textbooks, and compares student responses to these automatically created references.

To evaluate the system, we will use authentic student work collected from five different exam sessions across three undergraduate courses at the Faculty of Mathematics (two first-year courses and one fourth-year course) and two master's-level courses at the University of Arts. Student responses will be graded by both instructors and the platform, and the results compared to assess grading consistency and the quality of feedback. While a large-scale comparative study is planned as a future step, this evaluation aims to demonstrate the potential of the platform to provide instructors with a tool for creating, distributing, and evaluating assessments in an entirely online environment, streamlining the grading process while enabling high-quality AI-generated feedback.

To encourage wider use and collaboration, the platform will be made available as open-source software. Looking ahead, we plan to improve the tool by supporting more languages and adding features that explain how the AI comes to its decisions, so everyone can better understand and trust the results. Overall, this project shows how LLMs can play a practical role in modernizing educational assessment, providing scalable and efficient tools that meet the changing needs of both teachers and students.

Keywords: AI in higher education, automated grading, LLM, ChatGPT

References:

[1] Kooli, C., & Yusuf, N. (2025). Transforming educational assessment: Insights into the use of ChatGPT and large language models in grading. International Journal of Human–Computer Interaction, 41(5), 3388-3399.

[2] Jukiewicz, M. (2024). The future of grading programming assignments in education: The role of ChatGPT in automating the assessment and feedback process. Thinking Skills and Creativity, 52, 101522.

[3] Lundgren, M. (2024). Large Language Models in Student Assessment: Comparing ChatGPT and Human Graders. arXiv preprint arXiv:2406.16510.

The Challenges of AI Integration in Educational Systems

Vladimir Pavlović[1], Bogdan Novaković[2] Nataša Milosavljević[1], Olivera J. Bogunović[3]

[1]University of Belgrade, Faculty of Agriculture, Serbia, vlaver@agrif.bg.ac.rs, natasaglisovic@gmail.com
[2] University of Milano-Bicocca, Department of Human Sciences for Education "Riccardo Massa," Italy, b.novakovic@campus.unimib.it
[3] Harvard University, Harvard Medical School, USA, obogunovic@mclean.harvard.edu

The integration of artificial intelligence (AI) into education has evolved from early adaptive learning systems to advanced generative AI capable of producing original educational content. This technological shift is reshaping pedagogy, assessment, and administration. Large language models demonstrate significant potential in curriculum development, generating customized lesson plans, learning objectives, and assessments aligned with educational frameworks such as Bloom's Taxonomy. These tools offer educators time savings of 30-50% in content preparation, though output quality varies across disciplines. Multilingual content generation expands access to educational materials for diverse learners, though translation quality remains inconsistent for less commonly taught languages. In STEAM education, AI systems combine language processing with computational engines to create adaptive problem sets and programming exercises. The integration of diffusion models enables the development of multimodal learning materials, including interactive simulations and immersive historical recreations, which preliminary research suggests can improve student engagement by 20-35%.

Despite these advancements, AI implementation faces substantial challenges. Structured integration, bias mitigation, and preservation of human-centric learning remain essential for realizing benefits without compromising educational integrity. While standardized outputs generated by AI systems may inadvertently constrain pedagogical creativity and innovation, technical limitations include reliability concerns, with STEAM content exhibiting error rates of approximately 22%, and difficulties in handling higher-order thinking tasks, particularly in humanities disciplines requiring nuanced analysis.

The increasing adoption of AI in education introduces cognitive biases that affect both learning and teaching. It has been found that automation bias leads to uncritical acceptance of AI outputs, with students being 40% less likely to verify AI-generated information compared to traditional sources. Algorithmic anchoring limits exploratory thinking, reducing alternative problem-solving approaches by 28% in mathematics. Confirmation bias is amplified as adaptive systems prioritize content matching existing knowledge, while authority bias diminishes critical evaluation. AI also reshapes attention patterns. Microlearning modules increase task-switching by 35%, potentially impairing deep focus. Cognitive offloading through AI summarization reduces long-term retention by 18%, and multimodal environments elevate distractibility markers by 42%. Mitigation requires balanced AI use (30-50% integration shows optimal results), bias-aware design, and neuroadaptive systems that monitor engagement.

Ethical considerations present additional complexities. Analysis reveals persistent issues with bias amplification in generated content, while academic integrity frameworks struggle to accommodate AI-human collaboration. Data privacy concerns persist regarding the collection and use of sensitive student behavioral data through adaptive learning systems.

Systemic barriers further complicate adoption. Recent surveys indicate that 73% of educators lack adequate training to effectively evaluate AI-generated content, while infrastructure limitations prevent 88% of institutions in the Global South from utilizing GPU-intensive tools. The absence of standardized validation protocols for educational AI, unlike the rigorous review processes applied in fields like medical AI, leaves institutions without clear guidance for implementation. These challenges

risk exacerbating existing educational inequalities if not properly addressed. Successful integration of AI in education requires a balanced approach that combines technological innovation with pedagogical integrity. This includes developing comprehensive educator training programs, establishing robust evaluation frameworks, and maintaining a focus on human-centered educational values. The path forward must carefully navigate these challenges to realize AI's full potential while ensuring equitable access and preserving the quality of education.

Having in mind that in the context of STEAM (Science, Technology, Engineering, Arts, and Mathematics) education, AI can play a pivotal role in enhancing interdisciplinary learning and nurturing creative problem-solving skill, we sought to create an AI-assisted educational model that not only engages students but also fosters critical thinking and creativity in an interactive setting. Therefore we organized workshops in a form of a four-week educational program during which we introduced students to AI tools while encouraging critical engagement with digital technologies. The results we obtained through organization of these workshops based on creative use of AI in education and our comprehensive examination of educational AI based technologies identifies three critical areas of concern that emerge from current applications. First, technical limitations related to model interpretability and system reliability pose barriers to effective deployment. Second, ethical considerations regarding student data protection, algorithmic fairness, require careful navigation. Third, fundamental tensions arise between the efficiency gains offered by automation and the need to maintain human-centered educational values and practices. These challenges present barriers to effective implementation and risk exacerbating existing educational inequalities if not properly addressed.

Keywords: AI in Education, Ethical AI, Personalized Pedagogy, Algorithmic Bias, Human-AI Collaboration

References:

[1] Luckin, Rosemary & Cukurova, Mutlu & Kent, Carmel & Du Boulay, Benedict. (2022). Empowering educators to be AI-ready. Computers and Education: Artificial Intelligence. 3. 100076.

[2] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, Gjergji Kasneci, (2023) ChatGPT for good? On opportunities and challenges of large language models for education, Learning and Individual Differences, Vol 103, 102274,

[3] Vladimir Pavlović, Nataša Milosavljević, Vera Pavlović, Branislav Vlahović, (2023), Artificial intelligence and digital technologies in digitally-supported university education Mathematics for Human Flourishing in the Time of COVID-19 and Post COVID-19 Published by De Gruyter, 2023, 195-200

[4] Zawacki-Richter, O., Marín, V.I., Bond, M. *et al.* Systematic review of research on artificial intelligence applications in higher education – where are the educators?. *Int J Educ Technol High Educ* 16, 39 (2019).

[5] **UNESCO (2023).** *"AI and Education: Guidance for Policy-Makers."* **UNESCO Publishing**

A Decade of Student Voices: Leveraging LLMs to Scale Analysis of Student Feedback for Actionable Course Improvement

Irena Vodenska[1], Jelena Jovanovic[2], Vladan Devedzic[2,3]

[1] Administrative Sciences Department, Metropolitan College, Boston University, vodenska@bu.edu
[2] Faculty of Organizational Sciences, University of Belgrade, Belgrade, Serbia, jelena.jovanovic@fon.bg.ac.rs;
[3] Serbian Academy of Sciences and Arts, Kneza Mihaila 35, Belgrade, Serbia, devedzic@gmail.com

This paper analyzes student feedback for a specific course running in multiple distinct course sections per semester, for four semesters per year, Fall, Spring, and two Summer, spanning from 2013 to 2024. The aim is to apply large language models (LLMs) to systematically process open-ended student evaluations and identify actionable insights for educators. Specifically, we separate comments into several course-related and instructor-related domains, distinguishing between positive remarks and negative or improvement-oriented ones. Katz et al. (2024) show that generative text models can be effectively used to build qualitative codebooks for student evaluations of teaching, offering both scalability and consistency. Their work demonstrates that LLM-based coding produces valid categories that align closely with human-generated frameworks, while at the same time reducing the time and cost typically associated with manual thematic analysis. Moreover, Parker et al. (2025) demonstrate that when paired with effective prompting strategies, LLMs can achieve human-level performance in analyzing educational survey feedback, providing flexible support for both inductive (bottom-up) thematic analysis and deductive (top-down) classification, enabling a faster, more reliable, and rigorous interpretation of student feedback than traditional approaches.

Building on this work, our approach begins by using an LLM (GPT4o-mini) to decompose each student response into semantic units, which enables the capture of multiple impressions within a single response. For example, one student may simultaneously praise “the professor’s clarity and real-world examples” while noting “the heavy workload and unclear grading.” These units were also assigned polarity (positive / negative) based on the type of open-ended question they originated from: units that originated from responses asking about positive aspects of the course or instructor were labelled as positive, whereas those originating from questions asking about potential course improvements were labelled as negative. A significant portion of student responses originated from questions without clear polarity (i.e., those generally asking about student opinion about course and/or instructor), in which case the LLM was instructed to not only identify semantic units within a student response but also to label each unit based on its sentiment as positive, negative or neutral.

The extracted units are then classified by the LLM into one of the following topical categories: 1) Course Content; 2) Course Organization and Logistics; 3) Course value and relevance, 4) Materials and Resources; 5) Assignments and Grading; 6) Instructor’s Teaching Approach; 7) Instructor’s knowledge and expertise, 8) Instructor’s Clarity and Explanations; 9) Instructor’s Feedback and Support; 10) Instructor’s Demeanor and Communication; 11) Overall Impression, and 12) Other. These categories were previously inductively derived from the students’ responses by first asking the LLM to identify the main themes in a randomly sampled subset of student responses and defining the first set of categories based on those themes; and then, through several iterations of asking the LLM to classify randomly sampled subsets of responses and refining the categories based on the classification results.

After the semantic units extracted from student responses were classified into one of the above-listed categories, the next and final step in our analytics pipeline was summarization. In particular, we have prompted the LLM to summarize units from the same category, being of the same polarity (positive/negative) and originating from the same course edition. That way, for each course edition

and each category, we obtained a summary of students' positive and negative (i.e., improvement-oriented) opinions, as expressed through their responses on the open-ended survey questions. The main themes identified through summarizations highlighted instructor expertise, clarity, and practical applications of content, with remarks such as "Assignments mirrored real-world problems we might face at work." Improvement-oriented themes emphasized workload, grading transparency, and pacing, as illustrated by "Feedback came too late to be useful." To evaluate the classification accuracy, we used the GPT5-mini to apply the LLM-as-a-Judge approach (Zheng et al., 2023), where a superior LLM is prompted to assess the correctness of the results produced by another LLM, to evaluate the classification of semantic units extracted from the students' comments using GPT4o-mini. Considering the large number of classified semantic units (over 15K), we randomly sampled 50 positive and 50 negative units from each of the last 5 years (2021-2025), adding up to 500 units in total, and ran the LLM-as-a-judge process. We then manually examined the evaluation done by the "LLM judge" and noted where we agreed/disagreed with its judgment, as well as the ambiguous cases (e.g., multiple categories could have been assigned). Evaluation of each semantic unit by the "LLM judge" was checked by two researchers (paper co-authors), and any disagreements were resolved through discussion. We found that the LLM-as-a-judge approach was correct most of the time in evaluating the classification results. The errors that occurred could largely be attributed to the ambiguity of some of the students' comments and multiple categories being viable candidates for the same comment.

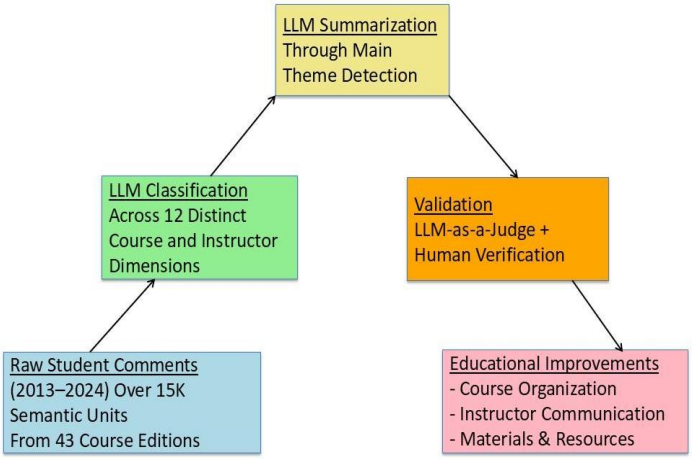


Figure 1 shows the pipeline for analyzing qualitative student feedback – raw comments are decomposed into semantic units, classified with LLMs, validated through LLM-as-a-Judge and humans, and distilled into actionable course improvements. Applying this framework across 43 semesters yielded 9,674 positive and 6,044 negative/improvement opinions. Longitudinal analysis revealed both continuities and shifts in student perceptions.

Fig. 1: Pipeline for analyzing student feedback

Praise for instructor clarity and expertise remained consistently strong, while concerns about course logistics, workload, and grading practices fluctuated with changes in course design.

Keywords: NLP, LLM-Assisted qualitative coding, Student course evaluations, LLM-as-a-judge

References:

[1] Katz, S., Doe, J., & Smith, M. (2024). Using generative text models to create qualitative codebooks for student evaluations of teaching. International Journal of Artificial Intelligence in Education. Advance online publication. https://doi.org/10.1007/s40593-024-00414-0

[2] Parker, M. J., Anderson, C., Stone, C., & Oh, Y. (2025). A large language model approach to educational survey feedback analysis. International Journal of Artificial Intelligence in Education, 35(3), 444–481. https://doi.org/10.1007/s40593-024-00414-0

[3] Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). Judging LLM-as-a-judge with MT-bench and Chatbot 0Arena. In Proceedings of the 37th International Conference on Neural Information Processing Systems (pp. 46595–46623). https://doi.org/10.5555/3666122.3669479

Classification of physics problems as a basis for the development of educational AI models

Milena Živković[1], Dubravka Živković[2], Marina Svičević[3], Aleksandar Milenković[3], Nemanja Vučićević[3]

[1]Department of Physcis, Faculty of Science, University of Kragujevac,Serbia, milena.zivkovic@pmf.kg.ac.rs
[2]Department of Psychology, Faculty of Philosophy, University of Niš, Serbia, dubravkazivkovic3@gmail.com
[3]Department of Mathematics and Informatics, Faculty of Science, University of Kragujevac,Serbia, marina.svicevic@pmf.kg.ac.rs, aleksandar.milenković@pmf.kg.ac.rs, nemanja.vucicevic@pmf.kg.ac.rs

The integration of artificial intelligence (AI) into education has opened new possibilities for personalized learning, adaptive assessment, and intelligent content delivery (Tapalova & Zhiyenbayeva, 2022). However, the effectiveness of such systems largely depends on the availability of high-quality, well-structured datasets that reflect the complexity of real educational content. In the domain of science education, and particularly physics, problem-solving tasks represent a core component of learning and assessment (Küchemann et al., 2024). Despite their importance, these tasks are rarely prepared in formats suitable for machine learning applications. To address this gap, we propose a structured classification framework for physics problems at the elementary school level (Shamshin, 2024; de Souza et al., 2024). Our approach focuses on the systematic annotation and organization of tasks with pedagogically and cognitively relevant features, creating a dataset suitable for training and evaluating AI models in education. By classifying physics problems based on problem type, cognitive complexity, physical quantities, and other key attributes, this framework supports the development of intelligent educational systems capable of adaptive task recommendation, personalized learning, and formative assessment. The problems were processed and annotated according to relevant criteria, including problem type (conceptual, quantitative, mixed), cognitive complexity level according to revised Bloom's taxonomy (Krathwohl, 2002), number of physical quantities and formulas, key concepts, and measurement units (Table 1). Problem complexity is determined by the number of reasoning steps required, from simple calculations (e.g., finding speed using $v = s/t$) to multi-concept problems (e.g., calculating acceleration while considering friction and inclined forces). For instance, determining a box's acceleration when pulled at an angle requires resolving forces, calculating friction, and applying Newton's laws - a 5-step analysis typical for 8th grade physics. The dataset covers key physics topics for grades 7 and 8, such as force and motion, oscillations, and optics.

Table 1. Classification Criteria for Physics Problems for AI Modeling

Category	Description
Problem Type	Quantitative, Conceptual, Mixed
Cognitive Level	Bloom's Taxonomy: Remember, Understand, Apply, Analyze
Physical Quantities	Number of physical quantities involved in the problem
Formulas	Number and complexity of formulas applied
Key Concepts	E.g., force, friction, reflection, refractive index

Units	SI units appearing in the problem
Presence of Figures	With figures, without figures
Problem Complexity	Number of steps required to reach the correct solution
Language	Serbian and English – for international use in AI applications

The classification was conducted with a focus on didactic clarity and was structured into a digital database suitable for training and evaluating AI models. This approach enables the development of systems for automatic problem classification and selection, content recommendation based on students' prior knowledge, and personalized learning. The proposed framework serves as a foundation for the development of adaptive and intelligent educational tools in physics teaching.

A prospective full-length paper should focus on the analysis and discussion of specific AI models and the ways in which these models could leverage the developed physics problem dataset. For example, transformer-based architecture could use this dataset for automated problem difficulty prediction by identifying patterns between problem features (e.g., formula count) and student performance metrics. Additionally, the dataset could be used to train reinforcement learning agents for personalized task recommendation systems that adapt the sequence of problems based on a student's mastery of individual concepts, such as force or optics. Furthermore, generative models could utilize the annotated problem structures to synthesize new, pedagogically sound physics questions while maintaining an appropriate level of complexity aligned with Bloom's taxonomy. Further extensions of this research could include comparing different Large Language Models (LLMs), such as o3 (OpenAI), Claude Opus 4 (Anthropic), DeepSeek R1, and Gemini Advanced (Google DeepMind), in the context of processing and applying this structured physics problem dataset. Such an analysis would help identify which architecture best support physics education tasks while also highlighting their limitations, such as the occurrence of hallucinations in derivations and reasoning processes. Of particular importance would be exploring the potential of fine-tuning open-source models on this specific dataset to achieve higher accuracy and reliability when solving similar types of physics problems. The results of such research could serve as valuable guidelines for the selection and adaptation of AI models in real-world educational scenarios.

Keywords: physics education, problem classification, artificial intelligence, Bloom's taxonomy, data set

References:

[1] Küchemann, S., Steinert, S., Kuhn, J., Avila, K., & Ruzika, S. (2024). Large language models—Valuable tools that require a sensitive integration into teaching and learning physics. *The physics teacher*, *62*(5), 400-402.

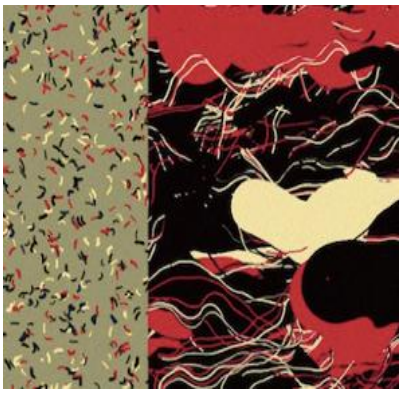
[2] Shamshin, A. (2024). Implementation of artificial intelligence in an online platform for solving and checking the solutions of physical problems. In *International Scientific-Practical Conference" Information Technology for Education, Science and Technics"* (pp. 400-417). Cham: Springer Nature Switzerland.

[3] de Souza, M. G., Won, M., Treagust, D., & Serrano, A. (2024). Visualising relativity: assessing high school students' understanding of complex physics concepts through AI-generated images. *Physics Education*, *59*(2), 025018.

[4] Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial intelligence in education: AIEd for personalised learning pathways. *Electronic Journal of e-Learning*, *20*(5), 639-653.

[4] Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. Theory into practice, 41(4), 212-218.

ARTIFICIAL INTELLIGENCE IN ETHICS, PRIVACY AND SOCIETY



Invisible Barriers: AI Bias and Cross-Border Privacy Risks in Automated Hiring

Jovan Kojić

Fakultet za preduzetnički biznis i menadžment nekretnina, Univerzitet "Union - Nikola Tesla", Beograd, Srbija, jkojictmf@gmail.com, ORCID: 0009-0004-5801-1199

Automated hiring technologies such as Applicant Tracking Systems (ATS) and AI-based screening platforms are rapidly reshaping recruitment. These systems promise increased efficiency, reduced costs, and improved alignment between candidate profiles and job requirements. Capable of processing thousands of applications within minutes, they significantly outperform manual methods in speed and scalability. However, beneath these operational advantages may be critical risks that challenge ethical, legal, and organizational norms. This paper explores how algorithmic bias and cross-border data privacy issues intersect in automated hiring, especially under the regulatory lens of frameworks like the General Data Protection Regulation (GDPR).

Despite their perceived objectivity, automated hiring systems are not free from human biases. ATS platforms often rely on structured inputs, keyword matching, resume formatting, and title alignment to filter candidates. As a result, qualified applicants using non-standard layouts, different terminology, or creative language may be discarded prematurely. These tools lack contextual understanding and adaptability, limiting their ability to fairly evaluate diverse applicant profiles.

AI-driven screening introduces a deeper layer of concern. Many such systems are trained on historical hiring data, which may embed patterns of systemic discrimination based on gender, ethnicity, age, or education. While developers often exclude protected characteristics, proxy variables such as employment gaps or school names can reintroduce bias. This creates an illusion of neutrality: decisions appear objective simply because they are made by algorithms. However, the lack of transparency, commonly referred to as the "black box" problem, makes it difficult for applicants to understand or contest decisions and for employers to audit these systems. This undermines not only fairness but also compliance with legal standards requiring explainability and human oversight.

The issue becomes more complex in multinational organizations where data flows are global. Hiring platforms frequently centralize operations via cloud-based systems that span multiple jurisdictions. As a result, applicant data might be collected in one country, analyzed in another, and stored in a third. While technically efficient, this decentralization introduces significant vulnerabilities. Data in transit is at risk of interception or corruption, particularly when encryption standards are uneven across systems or borders.

Legal challenges intensify in this cross-border context. The GDPR mandates that personal data transferred outside the EU must be protected to an "adequate" standard. Countries without equivalent protections, such as those with surveillance practices incompatible with GDPR principles, may face compliance risks. Even when companies use Standard Contractual Clauses (SCCs) or Binding Corporate Rules (BCRs) to justify transfers, maintaining full legal compliance remains a complex and resource-intensive task. Similar requirements exist in jurisdictions like Brazil (LGPD), Canada (PIPEDA), and South Africa (POPIA), creating a patchwork of obligations that global employers must navigate.

A particularly relevant GDPR provision is Article 22, which grants individuals the right not to be subject to a decision based solely on automated processing that significantly affects them. In hiring contexts, this means companies must either obtain explicit consent or ensure that meaningful human review is integrated into decision-making processes. These requirements are frequently overlooked in practice, especially when job seekers have no option but to submit to automated processes as part of an application.

Moreover, the lawful basis for processing data is another point of contention. While companies often cite “legitimate interest” as justification, the power imbalance in employment contexts makes genuine consent difficult to establish. Applicants typically cannot negotiate terms or opt out of automated screening without being excluded from consideration, challenging the fairness and legality of such data practices.

Without built-in mechanisms for accountability, bias correction, and privacy safeguards, automated hiring systems can reinforce discriminatory practices while exposing organizations to regulatory and reputational risks. The lack of international algorithmic governance further exacerbates the problem: a model trained in one legal environment may be deployed in another without adaptation, enabling systemic bias and legal blind spots to persist.

In conclusion, AI tools in hiring present a paradox. They offer organizations unmatched speed and cost savings but introduce risks that undermine fairness, inclusivity, and legal compliance. Ethical hiring in the digital era requires more than technical proficiency; it demands proactive governance. This includes regular audits for bias, transparent design processes, explainable outputs, and strict adherence to cross-border data protection laws. Without these safeguards, the invisible barriers embedded in code will continue to shape who gets hired and who gets excluded in ways that are neither fair nor lawful.

Keywords: Hiring, AI Bias, Data Protection, Cross-Border Compliance, Automated Decision-Making

References:

[1] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1–42. https://doi.org/10.1145/3236009

[2] Standard Contractual Clauses (SCC). (2021, June 4). European Commission. https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc_en

[3] Binding Corporate Rules (BCR). (n.d.). European Commission. https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/binding-corporate-rules-bcr_en#:~:text=Binding%20corporate%20rules%20(BCR)%20are,group%20of%20undertakings%20or%20enterprises

Artificial Intelligence and Citizen Identification through Matching Biometric Data with Audio and Video Recordings from Public Spaces: The Legal Status of Current Practice in Serbia

Dragan Manojlović[1], Ergin Hakić[2], Enver Međedović[3]

[1]Faculty of Law and Business, University MB Belgrade, Serbia,detore1914@gmail.com
[2]Department of Legal Sciences, Univesity of Novom Pazar, Serbia,e.hakic@uninp.edu.rs
[3]Department of Legal Sciences, Univesity of Novom Pazar,
Serbia,e.medjedovic@uninp.edu.rs

The identification of individuals in public spaces, based on audio and video surveillance recordings and their automated matching with stored biometric identification data (e.g., from national ID databases), through the use or misuse of artificial intelligence (AI) tools, is the subject of this research. The objective of this paper is to analyze the official connection established by state officials, within the scope of security operations, between the recordings of citizens generated in public spaces and the identification of individuals by automatically matching such recordings with biometric data stored in official identity document databases, utilizing AI technologies.

The study considers several examples where AI is used to identify individuals by extracting and matching visual or non-visual biometric features such as gender, date of birth, residence address, and other identifiers from national ID databases with footage captured in public environments. A commonly repeated statement in academic literature holds that state institutions may deploy AI technologies in ways that raise questions of oversight and accountability. To consolidate this control, public authorities and security services may authorize the deployment of advanced technological tools.

It is not uncommon to hear arguments that the security apparatus plays a central role in regulating societal interests. However, scientific analyses often refute such grandiose claims as unsubstantiated. This paper does not adopt a technophobic or conspiratorial stance toward AI use in public surveillance; rather, it critically examines the legal and normative frameworks or lack thereofthat govern such practices.

This paper questions why AI surveillance technologies have proven attractive to security agencies and institutions and emphasizes the need to approach their use with legal seriousness and analytical caution. The current unregulated deployment of AI tools for the biometric identification of individuals captured by public surveillance systems represents a legally and ethically challenging condition, particularly in jurisdictions where such systems operate without clear statutory authorization.

A core legal and ethical concern arises from the inability to selectively filter audio and video surveillance data matched with biometric datasets to support one purpose while excluding others, thereby amplifying the risk of misuse. The comprehensive profiling capabilities resulting from AI-enabled surveillance and biometric data fusion offer immense potential for both legitimate analysis and abusive or manipulative uses, often in the form of biased studies or targeted surveillance.

Despite arguments by state actors that these technologies are non-invasive and imperceptible to the general public, the absence of legal regulation raises fundamental questions about the lawfulness and legitimacy of their application in practice. The danger also lies in the normalization of such surveillance practices by state actors who do not recognize these AI-driven biometric matching processes as controversial or destabilizing.

The findings suggest that the AI-enabled linkage between ambient public surveillance and biometric identity data constitutes an unlawful data processing framework. This practice exposes citizens to heightened risks of misuse and lack of accountability.

Although the integration of biometric and surveillance data through AI is not a novelty in many European legal systems, most Balkan states lack a specific legal framework governing this practice. This paper highlights the need for a dedicated legal framework regulating the use of AI and emerging technologies in the field of citizen identification. The current fragmented legislative framework (a so-called “mosaic of laws”) creates legal ambiguity and enforcement inconsistencies.

Additionally, the paper proposes the establishment of a national oversight commission responsible for authorizing the use of AI in biometric identification, supervising database creation, safeguarding affected individuals’ rights, and overseeing the accountability of security institutions to prevent abuse. This body should have investigatory powers to monitor, restrict, and sanction improper use of such technologies.

Keywords: artificial intelligence, citizen identification, biometric data, audio and video surveillance, public space, legal framework in Serbia.

Application of AI for the Evaluation of Catalan-Based Cryptographic and Steganographic Methods

Muzafer Saračević, Suad Bećirović, Dženis Bajramović

University of Novi Pazar, muzafer.saracevic@uninp.edu.rs , s.becirovic@uninp.edu.rs , dz.bajramovic@uninp.edu.rs

This paper explores the application of artificial intelligence (AI) as a tool for testing and evaluating cryptographic and steganographic methods based on Catalan numbers and combinatorial problems. Building upon three prior methods [1,2,3] that introduced encryption data and data-hiding techniques based on Catalan objects, Dyck words, and Catalan random walks, this study leverages machine learning algorithms to analyze their security, robustness, and performance. AI-based testing frameworks are used to simulate attacks, detect hidden patterns, assess key sensitivity, and benchmark encryption strength under various conditions. The experimental results demonstrate that AI provides significant insights into the reliability and resilience of these approaches, offering a systematic and scalable method for validation and comparative analysis in security contexts.

In cryptography, several machine learning methods are used to distinguish two types of ciphertexts [2]. The basic questions of the analysis are: *Is it possible to identify the type of encryption method by machine learning models learned only from information in encrypted text? Are there significant differences between the previous encryption methods and the proposed Catalan-based method?* Our distinguishing attack approach includes: the ciphertext representation model analysis (features) and machine learning and testing of ciphertext-type classification models using several well-known machine learning methods.

In steganography, machine learning methods can be used to test the possibility of automated recognition of images containing steganographic content. Using machine learning from examples of original and modified images, it is generally possible to train a classifier that can recognize the class of changed images. One of the fundamental problems of learning such a classifier is the extraction of relevant features, i.e., making a description of an image, which is suitable for recognition. This can be accomplished by using separate image feature extraction methods, or by using machine learning methods that achieve that by themselves, for example, Deep Learning Neural Networks. Deep Convolutional Neural Networks (CNN) have several types of specialized layers intended to represent features from input data at different levels of abstraction [1].

Keywords: Artificial intelligence, Machine learning, Cryptography, Steganography, Testing.

References:

[1] Saračević M., Adamovic S., Miskovic V., Macek N., Sarac M. (2019), A novel approach to steganography based on the properties of Catalan numbers and Dyck words, Future Generation Computer Systems, Elsevier, Volume 100, Pages 186-197.

[2] Saračević M., Adamovic S., Miskovic V., M.Elhoseny, Macek N., M.M.Selim, K.Shankar (2021), Data Encryption for IoT Applications Based on Catalan Objects and Some Matchings Structures, IEEE Transactions on Reliability, 70(2): 819 – 830.

[3] Saračević M., Kumar Sharma S., Ahmad K. (2022), A Novel Block Encryption Method based on Catalan Random Walks, Multimedia Tools and Applications, Springer Nature, volume 81, pages 36667–36684

Challenges and Potentials in the Use of Artificial Intelligence for the Reconstruction and Simulation of Old Dialects and Language Varieties in the Balkans

Milena Šović[1], Milena Škobo[2], Marko Radojčić[3]

[1]Faculty of Economics and Engineering Management in Novi Sad - FIMEK, University Business Academy in Novi Sad, milena.sovich@gmail.com
[2]Sinergija University, Bosnia & Herzegovina, mskobo@sinergija.edu.ba
[3]Faculty of Civil Engineering, University of Belgrade, marko.p.radojcic@gmail.com

In an era of globalisation and digital standardisation, preserving linguistic diversity is becoming an increasingly urgent challenge. According to UNESCO estimates, a language or dialect disappears every two weeks, resulting in the irreversible loss of invaluable cultural and cognitive heritage (UNESCO, 2011). Dialects, archaic forms and chronolects are often inadequately documented, making them inaccessible to contemporary generations (Abbas et al., 2025). Although there are philological and ethnolinguistic initiatives to describe and archive these forms, traditional methods lack the capacity for dynamic speech simulation and integration into educational, cultural and digital systems. To what extent, then, can generative language models support us in this endeavour?

We will present an interdisciplinary framework for digitally restoring marginalised linguistic varieties, with a focus on applying contemporary artificial intelligence methods to the fields of dialectology and historical linguistics. A corpus-linguistic approach is adopted as a starting point, through which the extraction, normalisation and lexicostatistical refinement of historically relevant linguistic data are conducted. The theoretical framework distinguishes between archaic elements, temporally conditioned chronolects and individual linguistic systems (idiolects), placing particular emphasis on the concept of micro-innovation within small speech communities (Miličević Petrović, 2025; Karvalics, 2024). This layered variation aligns with Ranko Bugarski's (2003) insights into the sociolinguistic stratification of language in the Balkans, where dialectal diversity is shaped by complex historical, ethnic, and political dynamics. We will also analyse structural challenges, such as unstable orthographic conventions, lexical sensitivity to frequency and intra-dialectal variability, all of which complicate standardisation and digital modelling.

The technical section demonstrates how several ML and NLP techniques can be applied in data-limited conditions. We demonstrate how smaller LLM models (such as LLaMA, Falcon and Mistral) can be fine-tuned on micro-corpora to achieve satisfactory stylistic and grammatical adaptation. Retrieval-Augmented Generation (RAG) enables the integration of documented sources, such as archival dictionaries, grammars and transcripts, as an external memory during generation. We also consider methods for controlling style and lexicon, including pattern prompting, lexical filters and thematic response hierarchy.

An experimental digital humanities AI project has developed a multilingual chatbot that simulates the historical figure of Nikola Tesla. The system uses philologically refined vocabulary and controlled grammar to replicate Tesla's speech, and responds only in languages and styles from the late 19th to early 20th century. It provides answers to questions relevant only up to 1943, with the aim of not only informing users about Tesla's life, but also of authentically recreating his communication style within historical, sociolinguistic, and cognitive contexts.

The bot currently operates in four languages: Serbian, English, German and French. In Serbian and English, it responds as though addressing an educated audience. However, the English version is planned to be improved so that it consistently replicates Tesla's style as seen in his letters, interviews and autobiography. In German, the bot responds in the formal, academic style typical of intellectual circles in the Austro-Hungarian monarchy. In French, the bot's responses are processed at B2 level, in

line with Tesla's actual knowledge. It uses simple, grammatically correct sentences and avoids idioms and complicated stylistic figures.

Particularly inspiring for further development was the previous successful implementation of the Nušić chatbot, which was trained to imitate the style and humour of Branislav Nušić. In a controlled experiment, four bot excerpts were generated and compared with four authentic excerpts from lesser-known works by this writer. The evaluation was conducted by 50 teachers and professors of the Serbian language and literature, and the results showed that only 50% of the participants were able to accurately distinguish authentic from generated texts. This indicates a high success rate of style simulation, which opens the possibility of applying a similar methodology to the English language of Tesla's bot. The bot was tested in educational and museum environments, and its performance was evaluated based on user feedback.

Reviving extinct languages and historical dialects with the help of artificial intelligence poses many challenges, particularly when it comes to reconstructing accents and intonation without tone records. Unlike with the restoration of Hebrew, where the lexicon was deliberately modernised, AI must not introduce new words or structures independently in this context if we want to preserve historical fidelity. This raises the question of how modern users can interact with such a system, and vice versa, when they do not share the same linguistic references. Additionally, working with limited and non-standard corpora carries the risk that AI will 'invent' inauthentic language patterns. Therefore, strict control of the generation process is necessary, using filters, templates, and relevant sources. Human experts must remain the central authority throughout.

In the final, we explore the potential applications of such systems in language history teaching, digital humanities, virtual exhibitions and cultural diplomacy. Furthermore, we highlight ethical and methodological risks, such as creating 'authenticity without authentic speakers' the potential for generalisation and misuse of simulated identities, and issues surrounding cultural representation. Finally, we propose guidelines for future interdisciplinary projects, bringing together linguists, engineers and cultural experts to preserve intangible linguistic heritage in the digital sphere.

Keywords: Dialects, generative language models, archaic speech, RAG, cultural heritage.

References:

[1] Abbas, H. et al. (2025). AI-Powered Linguistics: The digital transformation of language and text analysis. *International Journal of Social Sciences, 3*(3), 11-25.
[2] Bugarski, R. (2010). *Language and identity*. Library 20th century. [In Serbian]
[3] Karvalics, L. Z. (2024). Artificial intelligence and language – A different approach to the problem. *Magyar Nyelvor, 148*, 589-600.
[4] Miličević Petrović, M. (2025). On 'linguistic boxes' and getting out of them: language data and linguistic analysis in contemporary language professions. In: S. Gudurić (Ed.). *Linguistic mosaic: applied linguistics in honor of Vesna Polovina*, 381-396. [In Serbian]
[5]UNESCO (2011). *Atlas of the World's Languages in Danger*. United Nations Educational, Scientific and Cultural Organization.

ARTIFICIAL INTELLIGENCE IN SECURITY

AI incidents and data integrity

Ana Kovačević

Faculty of Security Studies, University of Belgrade, Serbia, kana@fb.bg.ac.rs

Artificial intelligence (AI) has experienced widespread adoption across diverse sectors due to its capacity to enhance operational efficiency and economic competitiveness. However, the deployment of AI systems has also simultaneously introduced numerous security challenges and potential risks which demand careful consideration. As investments in AI development have increased substantially, corresponding investments in cybersecurity have become more critical. Ensuring the secure implementation of AI, particularly within critical infrastructure systems, necessitates the development of robust and resilient systems. Analysing real-world AI incidents provides valuable insights which may serve to enhance security mechanisms and prevent future vulnerabilities. Since the inception of artificial intelligence, researchers have identified various system vulnerabilities and associated risks. Promoting awareness of such potential AI hazards has proven instrumental in facilitating a deeper understanding of both the scope and severity of these risks. Moreover, such awareness provides a framework for developing AI tools which are not only more resilient, but also ethically sound.

It is essential to examine how these dynamics evolve in practical environments by systematically identifying such incidents, their underlying causes, and their consequent impacts, rather than relying solely on theoretical projections. In response to this urgent need, AI incident databases have emerged as crucial instruments for responsible AI development and governance. The primary objective of these initiatives is to methodically document and categorise incidents, thereby strengthening security measures, informing preventative strategies, and fostering transparency and accountability within AI systems management.

In addition to promoting transparency, such databases also facilitate enhanced collaboration between policymakers, researchers, and industry stakeholders. Within this context, the AI Incident Database (AIID) serves as a particularly significant platform for reporting and analysing AI-related incidents (AIID, 2025) in which AI technologies caused or almost caused harm. Similarly, recognising the unique security challenges posed by large language models (LLMs), the Open Web Application Security Project (OWASP) community developed the "Top 10 for LLM Applications", a resource which assists developers, security professionals, and organisations alike to understand and manage the most critical risks associated with LLMs (OWASP, 2024). This report identifies ten primary attack vectors targeting LLMs. Analysis of the AIID incident repository demonstrates that data integrity attacks emerge as the predominant threat vector in documented AI security incidents, with a marked escalation observed over the last three years. Analysis of incidents reported between January 1, 2023, and August 22, 2025, shows that integrity violations represent 63.3% of all cases, exceeding availability-related incidents (24%) and confidentiality breaches (12.7%) by far. The primary forms of integrity attacks include deepfake content synthesis, biased AI decision-making algorithms, and the systematic generation of false information. These findings are consistent with Schneier's observation (2025) that integrity-based attacks have become the predominant security threat in modern digital environments.

Large language models are fundamentally dependent on data quality, making the integrity of training data critically important. For AI system security, achieving the CIA triad (Confidentiality, Integrity, and Availability) is essential, where information security is realised when data confidential, maintains its integrity and remains available. Integrity ensures data is accurate and complete. Safeguarding data integrity is a matter of paramount importance, ensuring that data remains unaltered throughout its entire lifecycle without any unauthorised modification. Integrity violations occur either intentionally through malicious activities or unintentionally through errors. Maintaining total integrity is key for decision-making in sensitive domains such as medicine, law, and finance, as well as for maintaining

security between system components. In this context preserving integrity becomes particularly vital when AI agents are not only entrusted with responding to queries, but also making autonomous decisions (Schneier, 2025).

Although the phenomenon of fake content is not new, generative AI has created unprecedented capabilities for producing potentially deceptive content at extraordinary speed and scale, far exceeding human capacity. Inaccurate content can emerge at any stage of a model's lifecycle, posing particular risks for large language models due to their generative nature. This presents a particularly dangerous threat as it does not appear as an obvious error. Instead, the model gradually begins altering system behaviour, thus enabling the spread of inaccurate or biased content while making it difficult to detect the attack source. According to OWASP LLM classification, the creation of inaccurate content is primarily associated with the following types of attacks: prompt injection, data poisoning, and hallucination (OWASP, 2024). Prompt injection has been recognised as the most significant attack against large language model systems in recent years. This attack type exploits LLM vulnerabilities where models are unable to reliably distinguish between system instructions and user content. Data poisoning, on the other hand, aims to modify training data to introduce vulnerabilities, malicious code, or biases into models, thus compromising reliability, accuracy, and at times ethical output. The third category involves inaccurate information or hallucinations, caused by gaps in training data or biases resulting from imbalanced data representation.

Consistent with these findings, the "Bullshit Index" framework created by Liang et al. (2025) provides a systematic approach for identifying and measuring deceptive content generated by large language models. Their research demonstrates that incorporating Reinforcement Learning from Human Feedback (RLHF) and specific prompts (e.g. chain-of-thought) increases a model’s propensity for "paltering" and ambiguous language use. Notably, the bullshit index increased following RLHF implementation, indicating a reduced concern for truth accompanied by greater user satisfaction.

The integrity issue is not easily resolved, and as systems become increasingly complex, ensuring data integrity will be correspondingly more challenging. As Schneier emphasises, the fundamental question remains as to whether guaranteed integrity can ever be achieved. One possible solution to this problem, apart from multilayered technical measures throughout the entire data lifecycle, could lie in education and raising user awareness about potential issues. Educational institutions should adopt a more proactive and interactive approach to AI education in order to spark their students' interest by adapting curricula to current technological demands, thus preparing them for the ethical and successful navigation of the AI-driven world (Kovačević & Demić, 2024).

Keywords: AI incidents databases, integrity**,** Large Language Models, cyber security.

References:

[1] AIID (2025). AI Incident Database. Available at: https://incidentdatabase.ai/
[2] Kovačević, A., & Demić, E. (2024). The Impact of Gender, Seniority, Knowledge, and Interest on Attitudes to Artificial Intelligence. IEEE Access, 12, 129765–129775. doi.org/10.1109/ACCESS.2024.3454801
[3] Liang, K., Hu, H., Zhao, X., Song, D., Griffiths, T. L., & Fisac, J. F. (2025). Machine Bullshit: Characterizing the Emergent Disregard for Truth in Large Language Models (arXiv:2507.07484). arXiv. https://doi.org/10.48550/arXiv.2507.07484
[4] Ottenheimer, D., & Schneier, B. (n.d.). Data Integrity: The Key to Trust in AI Systems - IEEE Spectrum. https://spectrum.ieee.org/data-integrity
[5] OWASP (2024, November 18). OWASP Top 10 for LLM Applications 2025. OWASP Gen AI Security Project. Available at: https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/

An Approach for Security Support of Federated Machine Learning Against Backdoor Attack

Miodrag J. Mihaljević[1,2], Lianhai Wang[1] , Shujiang Xu[1] , Wei Shao[1]

[1]Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China

[2]Serbian Academy of Sciences and Arts, Belgrade, Serbia, miodragm@mi.sanu.ac.rs

Federated Learning (FL)

FL is a machine learning approach that enables training models across decentralized devices or servers while keeping data localized. This technique allows collaborative model training without sharing raw data, thereby addressing privacy concerns associated with centralized data storage. Each device or server computes model updates based on local data and shares updates only with a central server or aggregator, thereby ensuring data privacy and security. A key advantage of federated learning is its ability to leverage data from multiple sources without compromising individual privacy. By training models locally on user devices or servers, federated learning enables personalized model updates while preserving data confidentiality. This distributed learning approach is particularly beneficial in scenarios where data cannot be easily centralized, owing to privacy regulations, data sensitivity, or network constraints.

Backdoor Attack

Generally, the backdoor attack on FL can be performed in the following three steps.
1) Trigger Definition: The attacker secretly selects triggers in advance, typically choosing those with low-frequency characteristics that align with their specific objectives.
2) Poisoned Dataset Generation: The attacker selects a subset of the dataset, injects triggers into the samples, and modifies the labels to target labels. The training dataset is a combination of the clean and poisoned samples.
3) Model Backdoor Injection: The attacker uses a poisoned dataset and specific attack strategies. Overall, the attacker's objective is to modify the parameter of model θ to θ_p . The θ_p can be formulated as the following optimization problem:

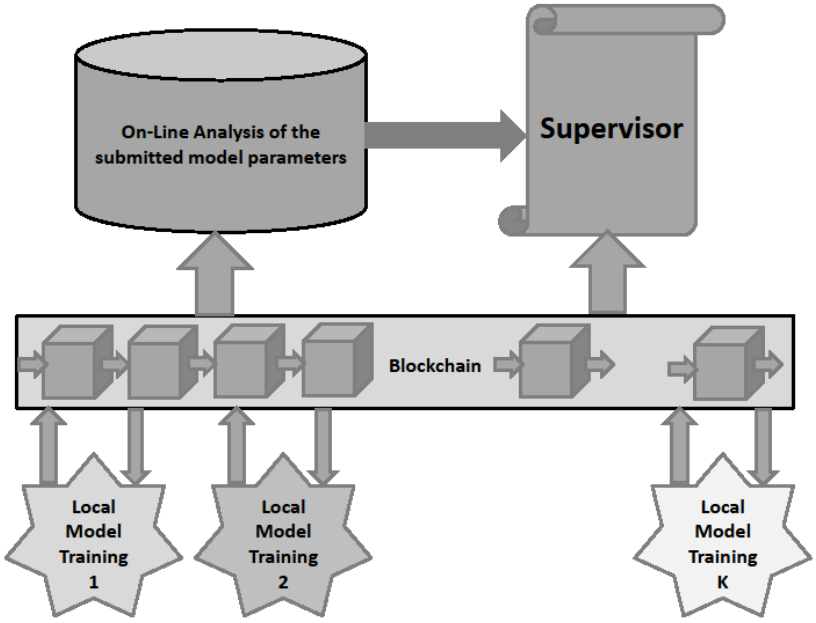
$$\theta_p = arg\ min_\theta [\sum_{(x_i,y_i)\in D_c} \mathcal{L}(f(x_i,\theta),y_i) + \sum_{(x*_j,y_i)\in D_p} \mathcal{L}(f(x^*_j,\theta),y_t)]$$

where $\mathcal{L}$ is the loss function, and D_c and D_p represent the clean training set and poisoned training set, respectively. $x^*_j = x_j \oplus \tau$ is a poisoned sample obtained by injecting a trigger τ into the clean sample x_j and y_t is a target output.

Proposed Architecture and Functionalities of Its Components

The main underlying idea is to build the security framework employing technical and administrative protection approaches. The technical approaches are based on employment of: (i) blockchain for recording data on learning process and the outcomes of the FL, i.e. the model parameters; and (ii) On-line AI-based analysis of the outcomes of local model training within the current and certain number of past training iterations in order to detect possible deviation of the parameters and, if any, issuing an alarm to the supervisor. The administrative protection is based on the contract based power of the supervisor to perform control over the local training processes.

Architecture of the proposed security framework is displayed in the following Figure.



Accordingly, the proposed FL model training is based on the following components.

-_*Entities for local training*. K entities that perform iterative local training of the model and in the i-th iteration generate updated model parameters using the parameters from the iteration i-1 as the initial ones.

- *Blockchain* that support the following main functionalities: (a) recording of the model parameters obtained from local training entities in each iteration, aggregation, updating and generation of the new model parameters; (b) platform for local training entities that has obligation to record into the blockchain signed hash value of the following concatenated data: (i) sample employed for the training, (ii) employed initial parameters of the model, and (iii) obtained updated parameters of the model;

- *Analyzer*. Component for the on-line analysis and detection of potential irregularities that is an AI based system for analysis of all the current locally trained model parameters submitted to the blockchain and a number of previous ones obtained from the local entities;

- *Supervisor*. Component for supervision of the entire FL training process. In order to provide Supervisor control of the training process, each entity that performs local model training is obliged to sign a contract that include statement on the Supervisor authorization to check the training data and the training process

A background for developing some of the above components are certain results reported in Chen at al. (2025), Wan et al. (2024), Nguyen et al. (2024), Papathanasopoulos et al. (2024).

Keywords: Federated learning, backdoor attacks, security, blockchain**.** .

References:

[1] Chen, C., Liao, T., Deng, X., Wu, Z., Huang S., and Zheng, Z., (2025, May/June). *Advances in Robust Federated Learning: A Survey With Heterogeneity Considerations*. IEEE Transactions on Big Data (vol. 11, no. 3, pp. 1548-1567).

[2] Wan, Y., Qu, Y., Ni, W., Xiang, Y., Gao, L., and Hossain E., (Third Quarter 2024) *Data and Model Poisoning Backdoor Attacks on Wireless Federated Learning, and the Defense Mechanisms: A Comprehensive Survey.* IEEE Communications Surveys & Tutorials (vol. 26, no. 3, pp. 1861-1897).

[3] Nguyen, D. H., Nguyen, P. L., Nguyeny, T. T., Phamz, H. H., and Tranx, D. A., (2024) *FedBlock: A Blockchain Approach to Federated Learning against Backdoor Attacks,* 2024 IEEE International Conference on Big Data (Big Data), Proceedings, (pp. 7981-7990).

[4] Papathanasopoulos, A., Apostolopoulos, P.A., Member, and Rahmat-Samii, Y. (2024, Jan.) Optimization Assisted by Neural Network-Based Machine Learning in Electromagnetic Applications. IEEE Trans. on Antennas and Propagation. (vol. 11, pp. 160-173)

Development of an Advanced Intrusion Detection System using Autoencoders

Marija Milošević[1], Vladimir Ćirić [1], Leonid Stoimenov [1], Ivan Milentijević [1]

[1]Faculty of Electronic Engineering, University of Niš, Serbia, marija.milosevic@elfak.ni.ac.rs, vladimir.ciric@elfak.ni.ac.rs, leonid.stoimenov@elfak.ni.ac.rs, ivan.milentijevic@elfak.ni.ac.rs

Introduction

In today's era of digital technologies, with increasing human dependence on them, where private and business data are stored on online locations, cyber-attacks are becoming a serious problem that is difficult to control. To address these threats effectively, development of advanced intrusion detection systems (IDS) is necessary. IDS is a crucial tool that allows identification and quick response to potential threats in real time. Traditional IDS systems, based on rules and heuristic methods, are often limited in their capability to recognize new and sophisticated attacks. AI is a powerful tool for improving IDS systems.

The Proposed System

This research focuses on the application of autoencoders (AE), which represent a promising method for anomaly and cyberthreat detection (Shone, 2018). The aim of this study is to explore the use of autoencoders in network intrusion detection systems by implementing attack-driven training, to create models that could differentiate and effectively detect the attacks. This extends our previous work (Ciric, 2024) by changing the context in which autoencoders are used.

Three referent, frequently used intrusion detection datasets have been chosen for training and evaluating the models - NSL-KDD (Tavallaee, 2009), UNSW-NB15 and CICIDS-2017 (Sharafaldin, 2018). Each of these datasets has a distinct set of attacks and features.

The proposed architecture is shown on Fig. 1. A dataset is chosen for autoencoder training and then adapted so it focuses on one attack. This means that a set with multiple classes is converted into a set of two classes - class containing the observed attack and a class containing all remaining traffic.

AE are a specialized type of neural networks designed for learning data representation. Their main function is to compress input data into lower dimensional latent representations and then reconstruct them back into their original shape. In the cyberattack detection context, autoencoders can be used for efficiently identifying deviations from data they were trained on (Mirsky, 2018). Four AE architectures have been used with variations in depth, types of layers, number of neurons in each layer and regularization functions. Three of them were described in (Ciric, 2024), having one, three, and five hidden layers, respectfully. In this paper, another architecture was implemented, having seven hidden layers. Since each dataset has a different set of features, the input layer of all autoencoders was based on the chosen input dataset and its features. Every architecture has been trained with each available attack from the datasets, regardless of the quantity of the attack, i.e. the number of available attack instances in the dataset. Having the roles in the training process reversed, the metrics that need to be observed are precision and recall for the attack the model was trained on, where recall is of greater importance because it highlights false negatives that are crucial for the system.

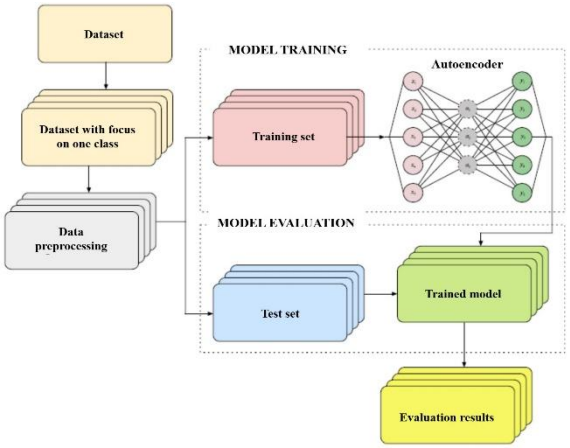


Fig. 1: System architecture.

Conclusion

Using autoencoders for specific attack detection could be a powerful tool in network intrusion combat, if used wisely. We have shown that models of different architectures can perfectly detect various types of attacks from referent datasets. This extends our previous work (Ciric, 2024) by changing the context in which autoencoders are used. Instead of training autoencoders the usual way, with normal traffic, and detecting malicious traffic as abnormal, the proposed methodology is for the models to "learn from the enemy", i.e. extract valuable patterns from attack representations and use the reconstruction error values to discern whether incoming traffic is that attack. The focus was on establishing whether the models can distinguish between a certain attack and the rest of the traffic that contains not only normal traffic but other types of attacks as well.

Acknowledgments: This work has been supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia [Grant Number: 451-03-137/2025-03/ 200102]. This work was also supported by the NATO Science for Peace and Security Programme funded project SENTAI (AI Sentinel: Enhancing Cybersecurity Through Artificial Intelligence), grant number G7593.

Keywords: Machine learning, autoencoders, cyber security, intrusion detection system

References:

[1] Shone, N. et al. (2018). A Deep Learning Approach to Network Intrusion Detection. *IEEE Transactions on Emerging Topics in Computational Intelligence* 2(1), pp. 41–50.

[2] Ciric,V., Milojkovic, A., Milosevic, M. (2024). Autoencoder-Based Network Intrusion Detection on Multiple Datasets. In: *2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON)*. pp. 1281–1286.

[3] Tavallaee, M. et al. (2009). A detailed analysis of the KDD CUP 99 data set. In: *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*. Pp. 1–6.

[4]Sharafaldin, I.; Habibi Lashkari, A.; Ghorbani, A.A. (2018). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. In: 4th International Conference on Information Systems Security and Privacy (ICISSP 2018). Pp. 108–116.

[5] Mirsky, Y. et al. (2018). Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. CoRR abs/1802.09089, arXiv: 1802.09089, http://arxiv.org/abs/1802. 09089.

Potential Applications of Artificial Intelligence for Countering Cyber-terrorism

Stefan Rasovic

Faculty of Security Studies , University of Belgrade, Serbia, stefanroyalcrowns@gmail.com

The Fourth Industrial Revolution has accelerated advances in artificial intelligence (AI), rendering it indispensable in domains ranging from automated data analytics and industrial process control to cybersecurity. The exponential growth of Internet-of-Things (IoT) devices and the increasing complexity of cyber threats underscore the critical need for AI solutions. This paper examines the emerging threat of cyber-terrorism and explores the potential of AI—particularly machine learning (ML) and deep learning (DL)—in detecting and mitigating cyber-terrorist activities aimed at critical infrastructure.

The beginning section briefly introduces foundational AI concepts, emphasizing ML and DL as dominant techniques today, combining supervised, unsupervised, reinforcement and emerging self-, semi- and transfer-learning paradigms. AI, now deeply embedded in analytics, facial recognition, anomaly detection, and cybersecurity, provides rapid and efficient decision-making capabilities. However, reliance on massive datasets and vulnerability to adversarial attacks present ongoing challenges (Russell and Norvig, 2021). Also, section assesses cyber-terrorism, highlighting definitional ambiguities and practical complexities. Despite no verified cyber-terrorist attacks meeting international legal thresholds to date, the expanding Internet of Things (IoT)—projected to reach 18.8 billion devices by 2024—and the increased cyber-physical integration through autonomous AI-enabled IoT controllers in critical infrastructure, considerably lower barriers for future cyber-enabled coercion and kinetic attacks. Concrete examples of potential kinetic attacks include disabling power grids, sabotaging transportation networks, or disrupting water supply systems, illustrating the tangible nature of cyber-terrorist threats. (Henschke, 2021).

The final section analysing the role and practical applications of AI in countering cyber-terrorism. Artificial intelligence (AI) has emerged as a crucial asset against increasingly sophisticated cyber threats, some of which may soon meet the threshold for cyber-terrorism as defined by international standards. A thorough conceptual reassessment indicates that cyber-terrorism can no longer be dismissed merely as hypothetical or theoretical. Primarily, this shift is driven by the proliferation of autonomous AI-enabled Internet-of-Things (IoT) devices embedded within Supervisory Control and Data Acquisition (SCADA), Distributed Control Systems (DCS), and other critical-infrastructure controllers, creating a tight coupling of cyber and physical domains (Putnik, 2022). These advanced systems allow digital disruptions to directly trigger real-world kinetic consequences, thus escalating the threat landscape dramatically. While no fully verified cyber-terrorist attack has occurred thus far, the expanding digital-physical attack surface clearly signals the increasing plausibility of such scenarios. Indeed, as the number of IoT devices surpasses tens of billions globally, many of these devices integrate AI modules that can act autonomously, optimize decision-making processes, and interact directly with physical actuators. Consequently, cyber-terrorists, equipped with comparable AI tools, could feasibly orchestrate highly coordinated attacks leading to widespread disruption or even physical destruction, targeting critical national infrastructure such as power grids, water systems, transportation networks, or healthcare facilities (Henschke, 2021).

To counteract these evolving threats, advanced machine learning (ML) and deep learning (DL) tools hold significant potential. Advanced machine- and deep-learning techniques can enhance defense by providing rapid image and face recognition, speech transcription, anomaly detection and classification of malicious network traffic, predictive intrusion monitoring and social-media surveillance of extremist content. By autonomously learning from vast data streams, AI systems can flag suspicious financial transactions, communication outliers and behavioural signatures indicative of cyber-terrorist preparation (Akilili, 2024). When integrated with SCADA and DCS, these tools enable real-time monitoring of both data payloads and metadata flows, identifying suspicious patterns indicative of intrusion attempts or malicious intent at unprecedented speeds and accuracy, far surpassing human analyst capabilities. In situations where cyber-terrorists specifically target AI-integrated IoT systems, advanced AI tools become even more vital. These scenarios demand predictive, proactive, and

resilient defences capable of rapidly adapting to novel attack strategies and techniques. Reinforcement learning models, for instance, can dynamically improve defensive actions by simulating various attack scenarios, allowing the AI systems to pre-emptively recognize vulnerabilities and automatically reconfigure security policies. Additionally, unsupervised and semi-supervised learning algorithms continuously identify emerging threats from previously unseen cyber-terrorist tactics, reducing detection latency and enhancing defensive resilience. AI-driven social network analytics and natural language processing algorithms further bolster defensive strategies by tracking extremist communications across encrypted platforms and the dark web (Nadjia, 2023. By uncovering hidden connections, intentions, affiliations, and emerging threats through real-time analysis of online discourse, security agencies can better predict and preempt cyber-terrorist actions targeting IoT-integrated critical infrastructures.

However, widespread deployment of these AI-driven security technologies must be matched with stringent regulatory frameworks and ethical standards. Challenges related to privacy violations, algorithmic bias, transparency, and accountability become even more acute given the intrusive capabilities of such powerful surveillance and analytics tools. Striking the right balance between AI-powered threat detection, human oversight, and protection of fundamental rights remains essential for sustaining public trust and the effectiveness of counter-terrorism measures.

Conclusion: AI is central to effective future cyber-defenses against increasingly plausible cyber-terrorism threats. AI-powered IoT devices embedded in critical infrastructure controllers tightly integrate cyberspace and the physical world, transforming cyber-terrorism from a theoretical possibility into a credible threat. Although no verified cyber-terrorist strike has yet occurred, the expanding attack surface signals that such events are plausible. A growing ecosystem of commercial AI platforms—deployed by corporations and government agencies—already outperforms traditional rule-based security tools. Continued technological advancement will further entrench AI at the centre of global cyber defense. However, successful mitigation will rely on balancing AI's predictive capabilities with robust ethical oversight, transparent governance frameworks, and strong collaboration among governmental agencies, the private sector, and academia to develop and deploy effective AI-driven cybersecurity solutions.

Keywords: artificial intelligence, machine learning, deep learning, cyberspace, cyber-terrorism, counter

References:

[1] Akilili, E. (2024). Artificial Intelligence in Counterterrorism: Navigating the Intersection of Security, Ethics and Privacy.
[2] Henschke, A. (2021). Terrorism and the Internet of Things: Cyber-Terrorism as an Emergent Threat. In: Henschke, A., Reed, A., Robbins, S., Miller, S. Counter-Terrorism, Ethics and Technology: Emerging Challenges at the Frontiers of Counter-Terrorism. Springer pp. 71-88.
[3] Nadjia, M. (2023). The role of artificial intelligence in combating cyber terrorism. IUS ET SCIENTIA, 9(2): 211-227.
[4] Russell, S., Norvig, P. (2021). Artificial Intelligence: A modern approach (4th edition). Pearson Education, Inc., London.
[5] Putnik, N. (2022). Sajber rat i sajber mir. Inovacioni centar Fakulteta bezbednosti. Beograd.

COMPUTER VISION

Efficient High-Fidelity Reconstruction of Dynamic Urban Scenes via Gaussian Splatting

***Aying*[1], *JianXun Cui*[1]**

School of Transportation Science and Engineering,Harbin Institute of Technology,China,
aying999g@163.com,cuijianxun@hit.edu.cn

Modeling dynamic 3D streets from images has significant practical applications across various domains, including city simulation, autonomous driving. Specifically, the digital twin of urban streets can serve as a simulation environment for autonomous vehicles, thereby significantly reducing training and testing costs. These application scenarios necessitate our ability to efficiently reconstruct 3D street models from captured data and render high-quality novel views in real-time.

With the development of neural scene representations[0], there have been some methods that attempt to reconstruct street scenes with neural radiance fields. To improve the modeling capability, EmerNeRF[0] introduces a self-supervised method for decomposing scenes into static and dynamic components, thereby improving the representation of highly dynamic environments. Although this strategy enables high-quality and realistic rendering of large-scale dynamic scenes, EmerNeRF requires excessively long training durations due to its extensive model parameters. Moreover, despite its significant success in static scene reconstruction, it struggles to effectively model dynamic elements, particularly fast-moving vehicles on the street, which are crucial aspects in autonomous driving environment simulation and limit its ability to fully simulate the highly dynamic and real-time nature of vehicular traffic.

3D Gaussian Splatting (3DGS)[0] has emerged as a powerful and efficient technique for novel view synthesis and 3D reconstruction, demonstrating significant potential for urban scene modeling. Its ability to represent complex geometries and textures with high fidelity while offering rapid rendering speeds makes it particularly appealing for large-scale outdoor environments[0].

We propose a novel explicit point-cloud based representation for dynamic scene reconstruction, enabling efficient recovery of 3D street environments from multi-view images. Concretely, we decompose urban street scenes into a static background and moving vehicles, each modeled independently with optimizable 3D Gaussian primitives. To capture vehicle dynamics, we represent each vehicle's geometry as a set of trackable point-cloud points—each storing learnable Gaussian parameters—and enforce temporal coherence via a pose-tracking network. Meanwhile, we employ a 4D spherical-harmonic (SH) appearance model driven by a 1D temporal function, dynamically predicting reflectance coefficients at any time step to faithfully represent time-varying lighting and material properties.

To enhance rendering speed and reduce training time, we adopted a more efficient rendering methodology within our 3DGS framework. First, we optimize the rendering pipeline to precisely localize Gaussians in the scene, enhancing rendering speed without altering visual fidelity. Second, we introduce a novel pruning technique and integrate it into the training pipeline, significantly reducing model size and training time while further raising rendering speed.

Comparative experiments were conducted on the Waymo Open Dataset against advanced methods such as EmerNeRF[0] and StreetGaussian[0], utilizing an NVIDIA RTX4090 hardware platform. Evaluating reconstruction quality comprehensively using PSNR, SSIM, and LPIPS, our proposed method demonstrated superior performance while achieving shorter training times and faster rendering speeds.

Keywords: 3D Reconstruction,Real-Time Rendering,Gaussian Splatting

References:

[1] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106.

[2] Yang, J., Ivanovic, B., Litany, O., Weng, X., Kim, S. W., Li, B., ... & Wang, Y. (2023). Emernerf: Emergent spatial-temporal scene decomposition via self-supervision. arXiv preprint arXiv:2311.02077.

[3] Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph., 42(4), 139-1.

[4] Liu, Y., Luo, C., Fan, L., Wang, N., Peng, J., & Zhang, Z. (2024, September). Citygaussian: Real-time high-quality large-scale scene rendering with gaussians. In European Conference on Computer Vision (pp. 265-282). Cham: Springer Nature Switzerland.

[5] Yan, Y., Lin, H., Zhou, C., Wang, W., Sun, H., Zhan, K., ... & Peng, S. (2024, September). Street gaussians: Modeling dynamic urban scenes with gaussian splatting. In European Conference on Computer Vision (pp. 156-173). Cham: Springer Nature Switzerland..

DrivingScene: A real-time and high-precision driving scene reconstruction method based on 2D Gaussian splatting

Qirui Hou[1], Jianxun Cui[1*]

[1]School of Transportation Science and Engineering, Harbin Institute of Technology, 24s032039@stu.hit.edu.cn
[2]School of Transportation Science and Engineering, Harbin Institute of Technology, cuijianxun@hit.edu.en

Over the past decades, early traditional methods based on multi-view stereo vision (MVS) were restricted by feature matching accuracy and computational efficiency, making it difficult to meet the real-time requirements of dynamic driving scenes. With the introduction of neural radiance fields (NERF)[1], which implicitly model complex scenes through ray tracing technology, significant breakthroughs have been achieved in the reconstruction quality of static scenes. However, due to its reliance on densely overlapping images and high computational costs, NERF cannot adapt to the rapidly changing street scenes and limited computational resources in autonomous driving scenarios. To address this issue, the 3D Gaussian splatting (3DGS)[2] technique emerged. This technique encodes scene details into explicit Gaussian primitives, significantly improving rendering speed and efficiency. Nevertheless, when dealing with complex dynamic street scenes, neither the original 3DGS technique nor subsequent improved methods that introduce time variables based on 3DGS can generate high-quality reconstruction results due to problems such as inaccurate object scale estimation and unreasonable Gaussian sphere constraints. For example, DrivingForward[3] uses a Gaussian neural network to generate Gaussian primitives, but due to improper multi-view consistency constraints, it produces a large number of artifacts in rendered images. Methods like MVSplat[4] , by ignoring the motion characteristics of vehicles and relying on highly overlapping image data, suffer from long model training cycles and convergence difficulties. DrivingScene introduces the 2DGS[5] renderer, which optimizes the multi-view rendering effect by converting the Gaussian ellipsoid into an elliptical disk, successfully eliminate the artifact problem caused by the perspective shift.

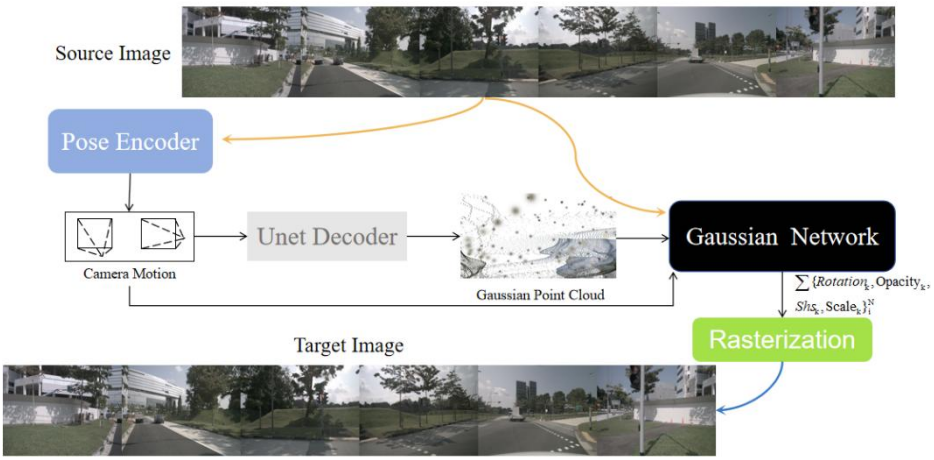


Fig1. The pipeline of DrivingScene

As illustrated in Fig1, our network is structured into three key components: a pose encoder responsible for predicting camera movement and orientation, a convolutional decoder that infers the scale and position of Gaussian primitives from multiple input panoramic images, and a module that takes the position of Gaussian primitives as input and outputs quaternion, scale, and spherical harmonic function information for primitive rasterization. During training, we adopt an end-to-end approach for the entire model, eliminating the need for additional modality information.

Drawing inspiration from the DrivingForward, our approach takes N sparse camera images as input, aiming to predict Gaussian primitives from the input view images. By predicting Gaussian primitives

on a per-pixel basis and assigning each pixel to the corresponding primitive, we determine the primitive positions, enabling feed-forward inference without test-time optimization. Accurate positioning of Gaussian primitives, which determines their centers, is critical for high-quality reconstruction. However, in driving scenarios, the limited overlap between sparse cameras constrains the multi-view geometric relationships.

The core of scale-aware Gaussian primitive localization lies in the pose encoder and convolutional decoder within our network architecture. During training, these components learn scale-aware depth from multi-frame surround views. During inference, they independently predict the true scale of Gaussian primitives from different surround views, achieving multi-view consistent scale-aware Gaussian localization. The pose network predicts vehicle motion, while the convolutional decoder extracts primitive scales from multi-view images. The key insight is to leverage the minimal overlap between spatially and temporally adjacent images for matching, which provides scale information and enables learning scale-aware camera motion and depth maps during training. We compute three photometric lossesL_{tm} 、 L_{sp}和L_{tm-sp} for each camera using spatial and temporal contexts. Additionally, considering the optimization of Gaussian primitives by the 2DGS renderer, we incorporate the depth distortion loss L_{depth} and the normal consistency loss L_{norm}. We also employ a smoothness loss L_{smooth} to encourage locally consistent depth outputs. The final loss function for scale-aware localization is:

$$L = L_{tm} + \lambda_{sp}L_{sp} + \lambda_{tm-sp}L_{tm-sp} + +\lambda_{depth}L_{depth} + \lambda_{norm}L_{norm} + \lambda_{smooth}L_{smooth} \quad (1)$$

Where λ represents the coefficients of each loss term during training.

And then, we propose a neural network which based 2D gaussian splatting capable of predicting detailed parameters of complex Gaussian primitives in aggregated driving scenes, such as quaternions, opacities and spherical harmonics. Using the previously obtained primitive scale information and camera pose matrices as inputs, the Gaussian network consists of an encoder, a feature fusion decoder, and four prediction heads for generating scaling factors, rotation quaternions, opacities, and spherical harmonics.

Keywords: Deep Learning, Neural Networks, Driving Scene Reconstruction

References:

[1] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106.

[2] Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph., 42(4), 139-1.

[3] Tian, Q., Tan, X., Xie, Y., & Ma, L. (2025, April). Drivingforward: Feed-forward 3d gaussian splatting for driving scene reconstruction from flexible surround-view input. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 7, pp. 7374-7382).

[4] Chen, Y., Xu, H., Zheng, C., Zhuang, B., Pollefeys, M., Geiger, A., ... & Cai, J. (2024, September). Mvsplat: Efficient 3d gaussian splatting from sparse multi-view images. In European Conference on Computer Vision (pp. 370-386). Cham: Springer Nature Switzerland.

[5] Huang, B., Yu, Z., Chen, A., Geiger, A., & Gao, S. (2024, July). 2d gaussian splatting for geometrically accurate radiance fields. In ACM SIGGRAPH 2024 conference papers (pp. 1-11).

Effects of Different Preprocessing Methods on Cell Segmentation using U-Net Neural Networks

Đorđe D. Nešković[1,2], Minja Vuković[1,2], Đurađ Budimir[1,3], Miloš Đorđević[2], Nadica Miljković[1,4]

[1]University of Belgrade - School of Electrical Engineering, Serbia
[2]Vinča Institute of Nuclear Sciences - National Institute of the Republic of Serbia, University of Belgrade, Serbia
[3]Wireless Communications Research Group, School of Computer Science and Eng., University of Westminster, London W1W 6UW, UK
[4]Faculty of Electrical Engineering, University of Ljubljana, Slovenia
e-mails: djordjeneskovic8@gmail.com, minja.vukovic00@gmail.com, d.budimir@westminster.ac.uk, mdjordjevic@vin.bg.ac.rs, nadica.miljkovic@etf.bg.ac.rs

1. Background: Microscopic image analysis is essential in biology and medicine for the detection, classification, and segmentation of specific structures. Automated approaches improve accuracy and reduce the subjectivity inherent in manual segmentation (Gamper et al., 2020). Although artificial intelligence and image processing for automatic cell recognition and characterization have been developed, their practical implementation is not yet entirely standardized. We aim to apply and evaluate a U-Net neural network for automatic cell segmentation of histopathological images (Ronneberger et al., 2015), with a focus on assessing the impact of preprocessing techniques on segmentation accuracy.

2. Methods: The software is implemented in Python version 3.11.0 (Python Software Foundation, Wilmington, DE, USA) within the Spyder Integrated Development Environment[1]. The PanNuke dataset used in this study is publicly available on GitHub and contains 7,901 semisynthetic RGB (Red, Green, and Blue) images (256×256 pixels) of various cell types (e.g., adrenal glands and colon), from both healthy and diseased tissues, with corresponding binary masks delineating cell contours (Gamper et al., 2020). Overall, 6,716 images are used for training and 1,185 for testing. Preprocessing involves three approaches. In the first case, image dimensions are scaled to 128×128 pixels to reduce data volume and accelerate training. Pixel values are normalized by dividing by 255, resulting in values ranging from 0 to 1. In the second approach, Principal Component Analysis (PCA) is applied to RGB components (Liu & Yang, 2008) to identify the two most informative components. Then, a new image is reconstructed and forwarded to the neural network for segmentation. In the third case, Independent Component Analysis (ICA) is applied to RGB channels (Liu & Yang, 2008). Histogram equalization is used for each Independent Component (IC) to enhance contrast, assuming that higher contrast improves segmentation. The IC with the highest contrast, assessed by the sum of normalized Root Mean Square (RMS) contrast, entropy, and Visual Information Fidelity (VIF) (Qureshi et al., 2017) is selected for training.

Apart from the IC with the highest contrast, the component with the second-highest contrast according to the contrast metrics is used for reconstructing the image (together with the IC with the highest contrast) that is further used to train the U-net. The implemented U-Net architecture follows the structure by Ronneberger et al. (2015), including number of layers, filter sizes, and skip connections. Otsu's adaptive thresholding is applied to the U-Net output to generate binary masks, producing pixel-wise segmentation probabilities. Fig. 1 shows an example of a segmented image. Segmentation performance is evaluated using Intersection Over Union (IOU) and Dice coefficient metrics (Wang et al., 2020). Five-fold cross-validation with shuffling is used for validation to ensure a pseudo-random distribution.

3. Results and Discussion: The segmentation results show that the performance obtained from the analysis of the rescaled and preprocessed images with the PCA method is very similar (IOU = 0.66 ± 0.17 and 0.65 ± 0.18, Dice = 0.81 ± 0.19 and 0.81 ± 0.20). This indicates that PCA manages to preserve information from the images while reducing noise and dimensionality, but without improvement compared to the results from the scaled data. On the other hand, the application of ICA followed by selection of a single "best" IC, according to the contrast-based metrics, resulted in

[1] Raybaut, P. (2009). Spyder-Documentation. Available at: pythonhosted.org.

noticeably lower segmentation performance (IOU = 0.37 ± 0.20, Dice = 0.51 ± 0.23), which only improve moderately after selecting additional IC for image reconstruction (IOU = 0.44 ± 0.22, Dice = 0.58 ± 0.26). The ICA-based approach is inferior to both the PCA-based and the scaled image approaches, indicating that information about the cellular structures is not contained solely in the ICs with the highest contrast. Although contrast may seem like a crucial feature to the human observer, the U-Net neural network likely relies on other image characteristics that are more relevant for successful segmentation. The results of the *k*-fold cross-validation (rescaled images: IOU = 0.64 ± 0.18 and Dice = 0.79 ± 0.21, ICA processed images (the "best" IC): IOU = 0.35 ± 0.21 and Dice = 0.50 ± 0.24, ICA processed images (the "best" two ICs): IOU = 0.43 ± 0.23 and Dice = 0.59 ± 0.25, PCA processed images: IOU = 0.63 ± 0.17 and Dice = 0.78 ± 0.19) show high consistency with the results on the test set, indicating good generalization of the U-Net model for segmentation. Similarly, the robustness of the U-net neural network depending on the input data fed to the U-net neural network was presented by Ronenberger et al. (the best IOU result on one set was 0.92, while on the other the IOU was 0.77). While our highest Dice coefficient is 0.81 is slightly better than those reported by more complex architectures like HoVer-Net (Dice = 0.80) by Gamper et al. (2020), it is important to note that their approach included both segmentation and classification across 19 cell types, whereas the proposed model focuses solely on segmentation. Also, our results should be interpreted with caution, given the fact that the reference binary masks may not be the most accurate ground-truth.

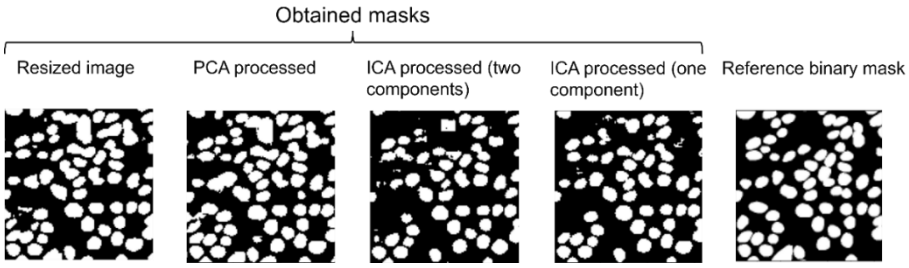


Fig. 1: Example of segmentation results for all three input data cases. The "ground truth" segmented regions from the adrenal gland tissue are shown in the far-right binary image[2] (Gamper et al., 2020).

4. Conclusion: We demonstrate that even standard U-Net architecture, combined with simple preprocessing steps such as PCA, can achieve competitive performance without the need for substantial computer resources. PCA preprocessing performs comparably to resized images, while ICA shows much lower segmentation performance. Our findings suggest that although preprocessing can assist in reducing noise and dimensionality, not all methods are suitable for enhancing model performance. In future work, we plan to expand the database with our own cell images. Following the segmentation process, the next step would be to assess method robustness on synthesized datasets and to adapt the methodology for calculating morphological characteristics of the segmented cells.

Keywords: cell segmentation, independent component analysis, principal component analysis, U-Net

References:

[1] Gamper, J., Koohbanani, N. A., Benes, K., Graham, S., Jahanifar, M., Khurram, S. A., & Rajpoot, N. (2020). Pannuke dataset extension, insights and baselines. arXiv preprint arXiv:2003.10778. https://doi.org/10.48550/arXiv.2003.10778.

[2] Liu, C., & Yang, J. (2008). ICA color space for pattern recognition. IEEE transactions on neural networks. https://doi.org/10.1109/TNN.2008.2005495.

[3] Qureshi, M. A., Beghdadi, A., & Deriche, M. (2017). Towards the design of a consistent image contrast enhancement evaluation measure. Signal Processing: Image Communication. https://doi.org/10.1016/j.image.2017.08.004.

[4] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference. https://doi.org/10.1007/978-3-319-24574-4_28.

[5] Wang, Z., Wang, E., & Zhu, Y. (2020). Image segmentation evaluation: a survey of methods. Artificial Intelligence Review. https://doi.org/10.1007/s10462-020-09830-9.

[2] PanNuke reference binary masks are shared within the database under open Creative Commons license. The database was firstly used by Gamper et al. (2020) and in the paper (Gamper et al., 2019) presented at the European Congress on Digital Pathology (https://doi.org/10.1007/978-3-030-23937-4_2)

Intelligent Syndrome differentiaton and prescription for Traditional Chinese Medicine

***Peiwu Qin*[1]**

[1]Institute of biopharmaceutics and Health Engineering, Tsinghua University Shenzhen International Graduate School, email: pwqin@sz.tsinghua.edu.cn

Traditional Chinese Medicine (TCM) is rooted in ancient philosophical frameworks and has been propagated over centuries in treating various illnesses and promoting overall wellness. At its core, TCM uses a four-pillar diagnosis ("inspection, smelling, inquiry, and palpation") to collect signs and symptoms to differentiate syndromes—n abstract concept describing a patient's holistic unhealthy conditions, akin to 'disease' in modern medicine (MM). However, the diagnostic process, particularly tongue diagnosis, heavily relies on the subjective judgment and personal experience of physicians, which can lead to variability and limit scalability. While systems incorporating MM parameters like retina fundus imaging or IR imaging are being explored, a comprehensive intelligent system for TCM diagnosis remains a challenge. Moreover, the scarcity of publicly available, high-quality TCM datasets severely hinders the performance of Multi-modal Large Language Models (MLLMs) in TCM diagnosis. To address these challenges, we introduce TCMPipe, a revolutionary framework that combines multi-modal language modeling and retrieval-augmented generation (RAG) to mimic the decision-making process of seasoned TCM physicians. First, we employ a novel tongue extractor that utilizes a zero-shot segmentation pipeline, enabling it to isolate the relevant tongue regions from an image without prior training or manual annotation. To generate tongue diagnoses, we then utilize the powerful visual comprehensive ability of a fine-tuned Qwen2-VL model to analyze the extracted tongue regions and produce detailed descriptions of its features. Finally, we introduce a prescription generator that integrates the patient's clinical metadata including chief complaint, history of illness, and pulse diagnosis with the AI-generated tongue diagnosis. This comprehensive input is then used to query a vast case bank, allowing the RAG-based system to formulate precise syndrome differentiations and prescriptions, thereby enhancing the automated diagnosis process. Furthermore, recognizing that standard metrics like ROUGE and BLEU fail to capture the nuances of TCM terminology, we designed two novel evaluation metrics based on TCM knowledge graphs to assess performance more accurately.TCM-metric A (Graph-based Shortest Path Evaluation) and TCM-metric B (Graph-based Token Similarity Evaluation), which compare predicted and reference diagnoses using a TCM knowledge graph.Extensive experiments were conducted on our proposed multi-center dataset, which comprises 1,317 tongue images and over 112,000 anonymized medical records.

TCMPipe-72B model achieved a ROUGE-1 score of 62.52% and a TCM-metric A score of 92.88%,far surpassing GPT-4o (15.23% ROUGE-1, 69.35% TCM-metric A) and other models like LLAMA 3.2

-Vision. This highlights our model's superior ability to understand and generate TCM-specific terminology. The results demonstrate that the diagnoses and prescriptions generated by TCMPipe significantly outperform current state-of-the-art methods, including GPT-4o, a finding that was also corroborated by a blinded user preference study with experienced TCM physicians.

As part of our commitment to fostering research in the field, this dataset will be continuously updated and made publicly available.

Kewords: TCM, Large language model, tongue image, image segmentation, herb prescription

References:

[1] Likun Zhang, **Peiwu Qin***, Xi Yuan, Jiansong Ji, Zhenglin Chen*. A 3.55-μm ultrathin, skin-like mechanoresponsive, compliant, and seamless ionic conductive electrode for epidermal electrophysiological signal acquisition and depression detection. **Exploration**, 2025.

[2] Zhicheng Du, Zhaotian Xie, Huazhang Ying, Likun Zhang, **Peiwu Qin**. Cognitive resilience: Unraveling the proficiency of image-captioning models to interpret masked visual content. **ICLR**, 2024

[3] Zhicheng Du, Zhaotian Xie, Yan Tong, **Peiwu Qin**. LAMPER: LanguAge Model and Prompt EngineeRing for zero-shot time series classification, **ICLR**, 2024.
[4] Dongliang Fan, Xi Yuan, Wenyu Wu, Renjie Zhu, Xin Yang, Yuxuan Liao, Yunteng Ma, Chufan Xiao, Cheng Chen, Changyue Liu, Hongqiang Wang**†**, **Peiwu Qin†**. Self-shrinking soft demoulding for complex high-aspect-ratio microchannels. **Nature Communications**, 2022.
[5] Muhammad Hassan, Hao Zhang, Ahmed Ameen Fateh, Shuyue Ma, Wen Liang, Dingqi Shang, Jiaming Deng, Ziheng Zhang, Tsz Kwan Lam, Ming Xu, Qiming Huang, Dongmei Yu, Canyang Zhang, Zhou You, Wei Pang, Chengmin‚g Yang**†**, **Peiwu Qin†**. Retinal disease projection conditioning by biological traits. **Complex & Intelligent Systems**, 2023.

Digitizing Serbian Periodicals: A Modular Pipeline for Historical Collections

Anđelka Zečević[1], Anastazia Žunić[1], Nikolina Vukša Popović[2], Božo Vuković[1], Lazar Velimirović[1]

[1]Mathematical Institute of the Serbian Academy of Sciences and Arts, Serbia, {andjelkaz, anastazia.zunic, lazar.velimirovic, bozov}@mi.sanu.ac.rs
[2]Faculty of Mathematics, University of Belgrade, Serbia, nikolina@matf.bg.ac.rs

Digital humanities is an interdisciplinary field that brings together traditional humanities and computer science to study cultural, historical, and societal data. Digitization of documents is important as it allows document preservation, wide accessibility, information retrieval, and additionally, it serves as a rich language resource. Numerous tools and platforms are being utilized in digital humanities to advance and enrich work with digitized documents. At the core of these tools are complex machine learning algorithms for image and text processing, trained on appropriately prepared datasets. Their most common functionalities include image quality enhancement, document layout analysis, optical character recognition (OCR), and post-OCR correction.

In this study, we share our experiences developing LibrAIfy, a software for digitizing old Serbian periodicals. Built on top of modern open-source tools and AI models, it is envisioned to support the transformation of digital scans to accurate, searchable text that can be incorporated in *Digitalna NBS*, the official platform of the National Library of Serbia dedicated to the exploration of digitized textual heritage. The development of LibrAIfy started with a corpus of scanned historical Serbian newspapers, collected by the National Library of Serbia, containing more than 400GB of data, with over 16,000 documents. The corpus includes four periodicals: *Vreme*, *Žena i svet*, *Amerikanski Srbobran*, and *Mali Žurnal*. These periodicals were published from the late 19th to mid-20th century and are characterized by a great diversity of graphic elements, non-standard formats, physical degradation, and lower-quality scans. To address the question of whether some of the widely used open-source OCR tools could be utilized to digitize these periodicals, we performed an explorative study of existing OCR tools, namely Calamari-OCR, docTr, Tesseract, and OCR4All. The evaluation revealed significant limitations, as these tools are typically built for linear, well-formatted, and more contemporary content. Consequently, their performance on historical periodicals was unsatisfactory, highlighting the need to develop a new OCR tool or use existing ones in a context more aware of document quality and layout structure.

Our pipeline consists of several modules. The first module processes the original input files, which are in PDF or WebBook format, and converts them into a set of images suitable for further processing. Since the newspapers and other periodicals are characterized by complex layouts, multiple column texts, various illustrations, images, and different typography across the pages, the second module performs document layout analysis. For document layout analysis, we use the YOLO-based object detection model (Redmon, 2016), which has been trained on a DocLayNet dataset (Pfitzmann, 2022). This allows the model to detect and identify layout elements such as titles, headings, text blocks, figures, and tables. This step is crucial as it allows the model to extract text regions from the document that will be further processed. In this way, the model will avoid the potential errors caused by non-textual elements, as illustrations. An example of document layout analysis is given in Fig. 1.

Given that the quality of the input document is often unsatisfactory, the extracted layout regions need to be denoised. To address this, the detected layout regions are passed through the image denoising module. This step is performed within our third module using Restormer model (Zamir, 2022), which has shown strong performance in image denoising tasks, fine-tuned on a ShabbyPages dataset (https://github.com/sparkfish/shabby-pages). This step improves the image quality of detected text regions and prepares them for more accurate OCR. The denoised text regions are further passed through the Tesseract OCR engine (Smith, 2007). The output from this module is searchable text.

However, by manual inspection, it is observed that the quality of text can still be improved. In order to enhance the output, we introduce the post-OCR correction module, which is designed to identify and correct the OCR errors. This module is based on a custom-made Long Short-Term Memory (LSTM) architecture trained on a SrpELTeC (Stanković, 2022), a dataset of 120 manually corrected and validated Serbian novels. The model is trained to correct the errors produced by the OCR module in order to improve the quality of the final text output. Finally, the pipeline ends with an output module that generates a standardized XML document in ALTO format that can be indexed and retrieved by the Digitalna NBS services.

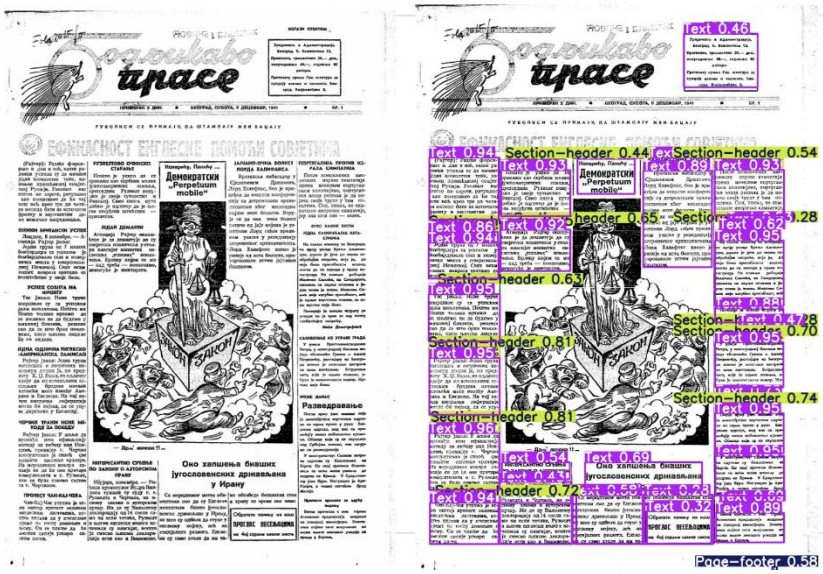


Fig 1. Document layout analysis with identified layout elements shown on the right.

In this work, we addressed the technical limitation of the existing OCR tools when applied to Serbian periodicals. By developing a modular pipeline designed for historical documents, we managed to close the gap in the digitization of underrepresented and topographically complex historical periodicals.

Acknowledgements. This work was supported by the Innovation Fund of the Republic of Serbia (Project IF ID: 53119) and the Ministry of Science, Technological Development, and Innovation of the Republic of Serbia through the Mathematical Institute SASA.

Keywords: computer vision, natural language processing, layout analysis, OCR, digital humanities

References:

[1] Pfitzmann, B., Auer, C., Dolfi, M., Nassar, A. S., & Staar, P. (2022, August). Doclaynet: A large human-annotated dataset for document-layout segmentation. In Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining (pp. 3743-3751).

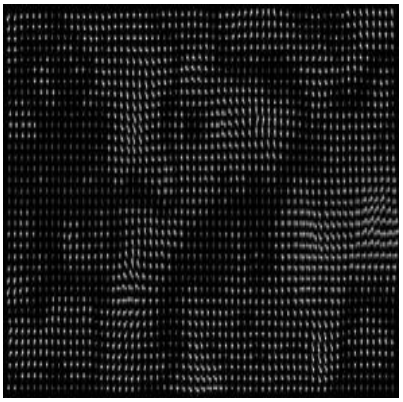
[2] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[3] Smith, R. (2007). An overview of the Tesseract OCR engine. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) (Vol. 2, pp. 629-633). IEEE.

[4] Stanković, R., Krstev, C., Todorović, B. Š., Vitas, D., Škorić, M., & Nešić, M. I. (2022). Distant reading in digital humanities: Case study on the Serbian part of the ELTEC collection. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 3337-3345).

[5] Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., & Yang, M. H. (2022). Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5728-5739).

DECISION MAKING AND PROBLEM SOLVING



Some possibilities to optimize federated learning processes

Mihailo Ilić[1], Mirjana Ivanović[1]

[1]Faculty of Sciences, University of Novi Sad, Serbia, milic@dmi.uns.ac.rs; mira@dmi.uns.ac.rs

Federated learning (FL) is a modern decentralized machine learning (ML) approach where multiple entities (edge nodes/clients) train a model collaboratively while keeping their data on-site, rather than sharing it in a centralized repository/storage. In this approach, clients train their own local models on locally stored data. These models are then collected by a central server (cloud), which uses different algorithms to aggregate them in order to create a global model i.e. each federated device shares its local model parameters instead of sharing the whole dataset which was used during training.

One important aspect of FL is the topology of the network of participants in the learning process. In a centralized topology, clients send their model parameters to a central server to produce a global model, which is in turn sent back to the edge nodes for further training. This is done numerous times, until convergence is reached, and the exact number of these global training steps depends on different aspects and the circumstances of the environment. Apart from the centralized topology, there also exist hierarchical and peer-to-peer approaches, where the topology may not even have a central server coordinating the learning process.
FL is a challenging research area and presents several compelling research problems.

- Communication efficiency: since FL involves frequent message passing, efficient communication and synchronization among FL entities is of significant importance.
- Privacy: there is a need for efficient implementation of privacy preserving technologies.
- Device heterogeneity: computing capacities of the FL entities are often heterogeneous, which makes it difficult to ensure efficient and fair learning.
- Data heterogeneity: local datasets can be highly heterogeneous in terms of quantity, quality, and diversity. Some edge nodes are more informative than others, and some may be underrepresented during model training.

Generally, there are two distinct approaches to collaborative ML: cross-silo and cross-device. Cross-silo FL involves collaboration between organizations where each "silo" holds its own data and model (like in healthcare or finance). Cross-device FL implies a huge network of devices (e.g. IoT devices) to train a model (personalized applications on individual devices).
To summarize, FL brings numerous security, privacy, communication, and computational benefits and a wide range of applications which can be categorized by various criteria (Kairouz, 2021). For critical systems (Dayan, 2021; Taïk, 2020), it is vital to obtain reliable models as soon as possible.

Our work focuses on organizational aspects of FL from two perspectives, particularly in environments with data availability and computational power constraints. The following research question is in focus: *How can the FL process be optimized during the setup phase to ensure optimal model performance and minimal resource usage?* The research plan includes two stages:
(1) An analysis of learning orchestration techniques to determine the FL strategy which yields the best models in cross-silo environments, it is crucial to obtain well performing models as soon as possible.
(2) Exploring the use of Watts-Strogatz (Stier, 2019) network priors in FL as an alternative to network pruning. The aim is to try to reduce computational load and communication overhead during model training at the edge. Such an approach could be very useful in IoT environments with limited computational power.

For the first stage it is interesting to explore efficient organization of the FL process, and we will analyze diverse strategies in cross-silo environments using different datasets. It will be interesting to

consider the relationship between data availability and the number of participating edge nodes for standard FL strategies: *incremental*, *cyclic-incremental*, *concurrent*, *semi-concurrent*. Some initial experiments with various data distributions between edge nodes were explored and it was concluded that incremental methods proved to be more robust in cross-silo settings with less data available. Concurrent methods caught up to incremental ones once more data became available in the federated environment.

We expect that critical systems based on cross-silo FL may benefit from initially using incremental methods, and later switching to concurrent methods of learning, once more data becomes available.

In the second stage we plan to concentrate on computational costs and communication overhead existing in massive environments like cross-device FL. Considering the computational cost of training deep models on IoT devices, we decide to use Watts-Strogatz (WS) network priors as a basis for hidden layer architectures in FL models, as it was done for centralized learning settings (Traub, 2023). This is a form of neural network sparsification done prior to the learning process. Network weights are pruned and rewired immediately upon the initialization of a new model, reducing the number of its weights. We expect that our experiments will show positive results like:

• selection of optimal WS priors leads to quality ML models;

• WS models can match the performance of regular, fully connected, models with the same number of training rounds in FL.

Different configurations in terms of sparsity and skip connections can be observed like in the Figure 1.

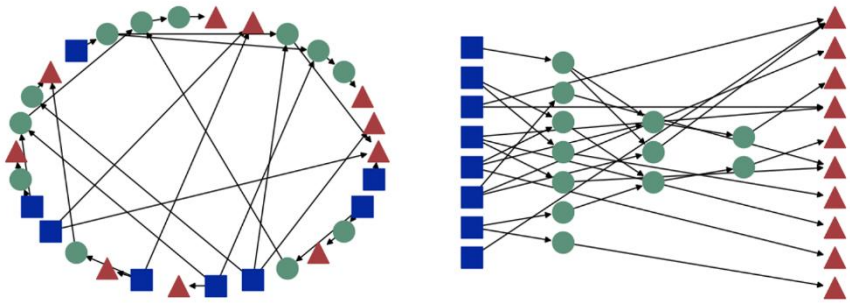


Fig. 1. The transformation of a Watts-Strogatz graph (left) into a hidden layer architecture (right). Blue squares have an in-degree equal to 0 and are placed in the first layer, red triangles have an out-degree of 0 and are placed in the last layer, while all other nodes (green circles) are in various layers in between

Keywords: Machine learning, Federated learning, Model Sparsification, Optimization

References:

[1] Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., ... & Li, Q. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature medicine*, *27*(10), 1735-1743.

[2] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and trends® in machine learning*, *14*(1–2), 1-210.

[3] Taïk, A., & Cherkaoui, S. (2020, June). Electrical load forecasting using edge computing and federated learning. In *ICC 2020-2020 IEEE international conference on communications (ICC)* (pp. 1-6). IEEE

[4] Traub, T., Nashouqu, M., & Gulyás, L. (2023, September). Efficient sparse networks from watts-strogatz network priors. In *International Conference on Computational Collective Intelligence* (pp. 163-175). Cham: Springer Nature Switzerland.

[5] Stier, J., & Granitzer, M. (2019). Structural analysis of sparse neural networks. *Procedia Computer Science*, *159*, 107-116.

Formalization of gyrovector spaces as models of hyperbolic geometry and special relativity

***Jelena Marković*[1]**

[1]Faculty of Mathematics, University of Belgrade, Serbia, jelena.markovic@matf,bg.ac.rs

Gyrovector spaces are algebraic structures that describe hyperbolic geometry in the same way that vector spaces describe Euclidean geometry. Although the operations of gyrovector spaces are neither commutative nor associative, the introduction of the notions of gyrocommutativity and gyroassociativity provides a stable framework within which fundamental theorems can be formulated in a form syntactically and structurally very close to their Euclidean analogues. This parallel not only facilitates the understanding of hyperbolic constructions but also opens possibilities for applying classical algebraic methods in a new, hyperbolic context.

Einstein's velocity addition is neither commutative nor associative, unlike the addition of velocities in classical mechanics. However, it is gyrocommutative and gyroassociative (in the sense in which these properties are defined, by introducing a specific correction factor, which in this case is not an abstract notion but an expression modeling relativistic effects—the so-called Thomas precession). This leads to the construction of Einstein's gyrovector space, which, as formally demonstrated in this work, satisfies Tarski's axioms together with the negation of the parallel axiom, i.e., it models hyperbolic geometry.

The subject of this work is the formal verification of the properties of gyrovector spaces and their generalization—normed gyrolinear spaces—within the Isabelle/HOL theorem prover. Through formal verification, one often discovers oversights in handwritten proofs—whether they are incorrect claims or statements that must be supplemented with additional assumptions in order to be valid—as well as new theorems. With the development of interactive theorem provers (Isabelle/HOL, Lean, Coq, etc.), formalization is becoming increasingly important in mathematics, with the aim of building a "digital library" of mathematics, where every lemma and theorem is checked down to the smallest detail. For young fields such as "gyromathematics," formalization brings confidence that future work will build on solid foundations, without relying on unchecked algebraic manipulations.

Particular attention in this work is devoted to the formalization of two central planar gyrovector spaces: Möbius gyrovector space, inspired by hyperbolic geometry, and Einstein gyrovector space, based on special relativity. In this work, their isomorphism has been not only mathematically proven but also fully formally verified in Isabelle/HOL, which represents a significant contribution, since such proofs had been omitted from the literature due to their complicated algebraic nature. In addition, the gyroisomorphism theorem has been completed and fully formalized, confirming that any structure isomorphic to a gyrovector space is itself a gyrovector space, thereby providing this field with a more complete and reliable theoretical foundation, as well as a more efficient method for determining whether a structure satisfies the axioms of a gyrovector space.

Furthermore, it has been formally proven that the Möbius gyrovector space is equivalent to the Poincaré disk model, while previous work has shown that the Poincaré disk model is a model of hyperbolic geometry. This means that the Möbius gyrovector space is also a model of hyperbolic geometry. In this way, a new formal model of the Poincaré disk has been developed within Isabelle/HOL, which is syntactically much simpler than classical approaches based on projective geometry, and is therefore more suitable for future use in the verification of more complex geometric theories. Finally, within this framework, gyrovector versions of well-known geometric results, such as the Pythagorean theorem and the law of cosines, have been reconstructed and formally verified, demonstrating the applicability of the developed formalism.

All of the above results have been published in joint work with Professor Filip Marić. In our previous work, we assumed that, in all gyrovector spaces under consideration, the set of norm values was embedded within the carrier set. This is not a strong assumption, as it is frequently the case. In our new work, however, we are no longer bound by this assumption. This change required us to shift our perspective on the representation of locales. Whereas previously we did not explicitly keep track of the domain, we now recognize the necessity of doing so. This approach to representation opens new avenues for exploration and formalization, including metric and topological properties.

Among the unpublished results, it is noteworthy that our work has established in Isabelle/HOL a solid foundation for the formal verification of the Mazur–Ulam theorem in the context of normed gyrolinear spaces. In the course of this development, an oversight was identified in one of the existing proofs of the theorem—a part that is potentially invalid or requires additional conditions in order to hold. Further directions for research include the extension to the three-dimensional setting, as well as the formalization of theorems concerning the relationship between gyrovector spaces and normed gyrolinear spaces, together with rigorous proofs of their various topological and metric properties.

Keywords: Gyrovector spaces, Hyperbolic geometry, Poincare Disc Model, Special relativity, Lorentz factor

References:

[1] Ungar, A. A. (2008). *Analytic hyperbolic geometry and Albert Einstein's special theory of* relativity. World Scientific. https://doi.org/10.1142/9789812772305

[2] Marković. J., Marić, F. (2025). *Formalization of gyrovector spaces as models of hyperbolic geometry and special relativity*. *Annals of Mathematics and Artificial Intelligence*, 1–35. https://doi.org/10.1007/s10472-025-09979-5

Hybridization of metaheuristics with machine learning methods

***Luka Matijević*[1]**

[1]Mathematical Institute of the Serbian Academy of Sciences and Arts, luka@mi.sanu.ac.rs

Optimization problems form the foundation of numerous business and industrial processes. Some of the most well-known examples include vehicle routing problems, task scheduling, resource allocation, and many others. Since finding optimal solutions to such problems is often computationally intensive, metaheuristic methods are widely used in practice to obtain sufficiently good solutions within a reasonable timeframe. A comprehensive overview of metaheuristic methods can be found in the book by *Talbi (2009)*.

On the other hand, machine learning has gained significant popularity in recent years, drawing growing attention from the research community, industry professionals, and the general public alike. Its ability to model complex patterns, adapt to data, and improve over time has opened new avenues in various fields, including optimization.

Recently, there has been a growing interest in combining metaheuristics with machine learning texhniues to leverage their complementary strengths. Specifically, certain machine learning techniques and concepts can be embedded within metaheuristic frameworks to guide the search process, adapt parameters dynamically, or learn from past search experiences. These hybrid methods, which blend metaheuristics with learning mechanisms, are commonly referred to as learnheuristics. A detailed overview of learnheuristics is presented in the paper by *Calvet et al (2017)*.

In this lecture, we will explore various ways in which machine learning techniques can be integrated into metaheuristic algorithms to enhance their performance and improve solution quality. Our primary focus will be on the *General Variable Neighborhood Search* (**GVNS**) method, which will serve as the foundational metaheuristic for illustrating these hybridization strategies. GVNS is a well-established and flexible method that allows for systematic neighborhood changes during the search process, making it a suitable platform for incorporating learning-based enhancements. Further details about GVNS can be found in the paper by *Hansen et al (2017)*. While GVNS will be the main algorithm under consideration, the techniques discussed, ranging from parameter tuning and adaptive decision-making to learning-based neighborhood selection, are not limited to this framework. They can be generalized and applied across a wide range of metaheuristic methods.

Furthermore, many metaheuristic methods incorporate some form of local search to improve solution quality by exploring the neighborhood of a given solution. A wide range of local search strategies exists—some of which are detailed in *Matijević (2025)*—and many of them stand to benefit from the integration of artificial intelligence techniques.

In this context, we will focus primarily on the incorporation of reinforcement learning into V*ariable Neighborhood Descent* (**VND**). VND is a deterministic variant of *Variable Neighborhood Search* (**VNS**) and is frequently employed as a local search component, especially within the GVNS framework. Unlike standard VNS, which incorporates stochastic elements, VND systematically explores predefined neighborhood structures in a fixed sequence.

The choice and ordering of these neighborhood structures play a crucial role in the efficiency and effectiveness of the search process. Poorly chosen sequences can lead to slow convergence or getting trapped in local optima. To address this, we aim to investigate the potential of machine learning, particularly reinforcement learning, to dynamically adapt and select neighborhood structures based on feedback from the search process.

Finally, we explore how ML techniques can be leveraged during the preprocessing phase—prior to the application of metaheuristics—to improve overall results. ML is frequently employed for tasks such as hyperparameter tuning and hyperparameter analysis, both of which significantly influence the performance of metaheuristic algorithms. Additionally, ML techniques have found use in problem-specific preprocessing strategies. For instance, in the context of the Vehicle Routing Problem, customer locations can be clustered before applying the metaheuristic, with each cluster assigned to a specific vehicle. While this approach may not yield an optimal solution, it can provide a strong initial solution that facilitates more effective and efficient subsequent search.

The research presented in this abstract is part of ongoing work conducted within the MI SANU Student Summer Internship program. Any results and insights obtained during the internship will be presented at this conference.

Keywords: Learnheuristics, Variable Neighborhood Search, Genetic Algorithm, Reinforcement Learning

References:

[1] Calvet, L., Armas, J.D., Masip, D. and Juan, A.A., 2017. Learnheuristics: hybridizing metaheuristics with machine learning for optimization with dynamic inputs. Open Mathematics, 15(1), pp.261-280.
[2] Hansen, P., Mladenović, N., Todosijević, R. and Hanafi, S., 2017. Variable neighborhood search: basics and variants. EURO Journal on Computational Optimization, 5(3), pp.423-454.
[3] Talbi, E.G., 2009. Metaheuristics: from design to implementation. John Wiley & Sons.
[4] Matijević, L., 2025. Metaheuristic approaches to the green vehicle routing problem including alternative fuel vehicles. Ph.D. thesis. Faculty of Technical Sciences, University of Novi Sad.

On a Machine Learning Approach for the Minimums Evaluation of Complex Multidimensional Functions

Nevena Mijajlović[1], Ajlan Zajmović[1] , Miodrag J. Mihaljević[2,3], Lianhai Wang[2] , Shujiang Xu[2] , Wei Shao[2]

1. Faculty of Science and Mathematics, University of Montenegro, Montenegro, nevenami@ucg.ac.me
2. Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
3. Serbian Academy of Sciences and Arts, Belgrade, Serbia, miodragm@mi.sanu.ac.rs

Introduction: We consider the multidimensional functions that perform mapping {R}n -> R known only in certain points (reduced table definition) and address the problem of minimums evaluation under certain constrains. A motivation for consideration of this problem appears from certain optimization issues regarding blockchain technology, and in particular regarding optimization of the pool mining (see for example Yu et al. 2024). Following Papathanasopoulos et al. 2024, Gajbhiye et al. 2025, and Wang et al. 2025, we employ machine learning based on neural networks training for addressing the considered problem. The talk reports the algorithm for finding the function local/global minimums and illustrative numerical examples.

Algorithm: We consider a variant of the approach known as “surrogate-assisted optimization” that replaces the direct optimization of an objective function with its surrogate model representation. The surrogate-assisted optimization algorithm include offline training phase prior to the optimization. The offline training aims to build an accurate model of the objective function and the effectiveness of the optimization depends on the initial training set size.

In order to evaluate effectiveness and correctness of the surrogate-assisted optimization, we first create a synthetic dataset by sampling points from a benchmark function within a predefined range of values. This dataset is then used to train a feedforward fully connected neural network (multilayer perceptron) so that it can learn the relationship between the input variables and the function’s output value.

We optimize the network using the AdamW optimizer with cosine annealing learning rate scheduling, along with a small weight decay to improve stability. For the loss function, we use Smooth L1 loss which is robust to outliers but still accurate near the optimum. Training is stopped early if the validation loss stops improving, which helps prevent overfitting.

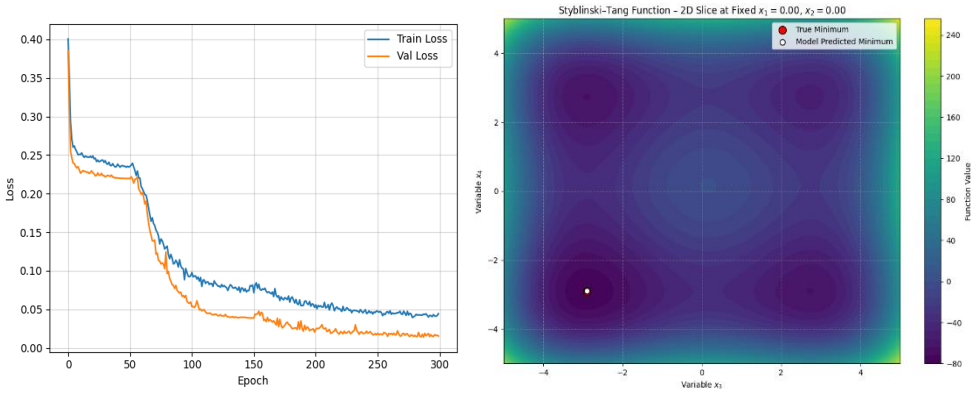
After training, the neural network acts as a surrogate model, replacing the original benchmark function in the optimization process. This makes optimization faster because the neural network is much quicker to evaluate than the original function. The search for the function’s minimum is then done using the Differential Evolution algorithm, which can find the global minimum efficiently.

This framework can be applied not only to well-known benchmark functions (Ackley, Styblinski–Tang, Rastrigin, etc., see Yang, 2010, for example) but also to real-world problems where evaluating the objective function directly would take too much computation time.

Illustrative Numerical Examples: In our numerical examples, we focus on a 4D conditional minimization problem, where the first two coordinates are fixed, and the optimization is performed over the remaining two variables. However, the network structure can be adjusted to handle problems with more or fewer dimensions depending on the task. For the 4D conditional optimization problem, we used a neural network with three hidden layers containing 128, 64, and 32 neurons, respectively.

Each layer uses the GELU activation function, and dropout regularization is applied after each hidden layer to reduce overfitting. The network outputs a single number, representing the predicted function value

As the first numerical illustration, we consider the 4-dimensional Styblinski–Tang function as the benchmark objective. The first two variables, x_1, x_2 are fixed at zero, and optimization is performed over x_3, x_4. Dropout rate is set to 0.05 to mitigate overfitting. A synthetic dataset of 3,000 points is generated by uniform sampling in the range [-5,5] for each dimension, and then split into 80% training and 20% validation sets. In this case, the predicted minimum (x_3, x_4)≈(-2.9010,-2.8957) yields a real function value of approximately -78.3312, differing from the true conditional minimum -78.3323, which was found in the point (x_3, x_4)≈(-2.90353398 -2.90353399) by only 0.0012.



A similar numerical experiment was performed on the 4-dimensional Ackley function with the first two coordinates fixed at 0.5 and -1.5 respectively, using a dropout rate of 0.1 and a dataset of 1000 points, split in the same way as for the previous example. The true conditional minimum was found at (x_3, x_4)≈(1.06×10−8,1.61×10−8) with a value of 4.6432, while the surrogate model predicted the minimum at (x_3, x_4)≈(−0.0345,0.0141), giving a real function value of 4.6508, only 0.0075 higher than the true minimum.

The results show that neural networks can be powerful surrogate models for global optimization tasks.

Keywords: Machine learning, neural networks, algorithms, minimums evaluation, numerical examples

References:

[1] Yu, Z., Chang, Z., Wang, L., and Min, G. (2024, Nov./Dec.). *Contract-Based Incentive Design for Resource Allocation in Edge Computing-Based Blockchain*. IEEE Trans. on Network Sci. and Engineering (vol. 11, pp. 6143-6156)

[2] Papathanasopoulos, A., Apostolopoulos, P.A., Member, and Rahmat-Samii, Y. (2024, Jan.) *Optimization Assisted by Neural Network-Based Machine Learning in Electromagnetic Applications*. IEEE Trans. on Antennas and Propagation. (vol. 11, pp. 160-173)

[3] Gajbhiye, P.A., Singh, S.P., and Sharma, M.K. (2025, Jun). *A comprehensive review of AI and machine learning techniques in antenna design optimization and measurement,* Discover Electronics, https://doi.org/10.1007/s44291-025-00084-9.

[4] Wang, Q., Wang, L., Xu, S., Zhang, Z., Shao, W., and Mihaljevic., M.J. (2025, Jan.) *Single-Layer Trainable Neural Network for Secure Inference*. IEEE Internet of Things Journal (vol. 12, pp. 2968-2978)

[5] Yang x.-S, *Test Problems in Optimization*, X.-S. Yang, Ed. Hoboken,NJ, USA: Wiley, 2010.

Metaheuristic Clustering of Incomplete Data: Towards Adaptive Objective Functions and Prior-Guided Initialization

Nataša Milosavljević[1], Tatjana Davidović[2]

[1]Department of Mathematics and Physics, Faculty of Agriculture, University of Belgrade,Serbia, natasam@agrif.bg.ac.rs
[2]Mathematical Institute, Serbian Academy of Sciences and Arts, tanjad@mi.sanu.ac.rs

Missing data is a pervasive challenge in domains such as medicine (Goel et al., 2024.), agriculture, and environmental monitoring. Traditional approaches either discard incomplete records or apply static imputations, which often introduce bias or obscure underlying patterns. Although neural network-based imputers are powerful, they typically lack transparency and require large datasets to perform effectively (Singh et al., 2024.).

Clustering algorithms form the backbone of exploratory data analysis and knowledge discovery. However, their effectiveness is often diminished in the presence of incomplete data (Saini et al., 2024.) — a frequent issue in practical scenarios such as medical diagnostics, remote sensing, and agro-environmental datasets. Traditional clustering techniques either ignore missingness, apply naive imputations, or rely on complete-case analysis, often resulting in biased or suboptimal outcomes. To overcome these limitations, we introduce a metaheuristic-based clustering paradigm that natively supports partial data and facilitates adaptive refinement.

In this study, we propose a metaheuristic clustering framework based on variants of Bee Colony Optimization (BCO) and Variable Neighborhood Search (VNS), specifically adapted for directly handling missing data. The method operates using a weighted objective function that minimizes intra-cluster distances.

Rather than treating missing data as a preprocessing issue, our approach incorporates it directly into the optimization objective. This enables reasoning under partial observability while leveraging adaptive attribute weighting and interpretable distance metrics. As such, the framework offers a principled alternative for applications where explainability, robustness, and flexibility are essential.

Our current implementation uses a customized VNS/BCO algorithm applied to the k-median clustering model (Davidovic et al., 2018.). Pairwise distances are computed as weighted sums of normalized, attribute-wise differences, with static imputations used as a fallback. The algorithm performs multiple restarts and local searches, optimizing candidate centroids by minimizing intra-cluster dissimilarity. This design ensures deterministic behavior and strong performance even in small or noisy datasets.

Building upon this foundation, we propose two key extensions to improve optimization quality and adaptability. First, we introduce **prior-informed initialization**, where distributions of historically successful centroid configurations are learned and used to bias the starting state. Such priors can be obtained from synthetic simulations, expert knowledge, or self-adaptive training and are expected to reduce convergence time while increasing solution stability.

Second, we introduce an adaptive weighting scheme that replaces fixed attribute weights with context-sensitive values, learned online during optimization. These weights are continuously updated based on each attribute's contribution to intra-cluster compactness, enabling dynamic re-evaluation of feature importance throughout the search process. This unified mechanism transforms the objective function into a dynamic landscape, introduces self-regulation, and improves robustness when clustering heterogeneous data, particularly in scenarios where feature significance may vary across clusters or iterations. The learning process can utilize gradient-free techniques, reinforcement heuristics, or feedback-driven meta-learning to maintain both global coherence and local adaptability.

Importantly, we highlight that metaheuristic methods such as VNS/BCO—even without embedded learning modules—can match or sometimes even outperform many state-of-the-art AI and machine learning algorithms, including KMeans++, DBSCAN, and autoencoder-based clustering models (Davidovic et al., 2018.). These advantages are further enhanced when the algorithm is augmented with prior knowledge and adaptive mechanisms (Capariño, 2024.).

To validate the proposed framework, we conducted initial experiments using benchmark datasets with varying degrees of incompleteness. Preliminary results show that our adaptive VNS/BCO-based clustering model outperforms baseline methods in both clustering accuracy and robustness under increasing proportions of missing data. These results demonstrate the benefit of combining adaptive weighting and prior-guided initialization, particularly in heterogeneous or noisy environments. Its advantage is the possibility to adapt its methodology by learning from both a relatively small initial dataset and newly arrived data.

In conclusion, the proposed framework combines the exploratory strength of metaheuristics with adaptive learning and domain-driven priors, yielding a robust and general-purpose clustering methodology. Future work will investigate the influence of missing data volume on clustering accuracy, deeper integration of feedback-driven adaptation, and further validation across diverse real-world domains.

Keywords: metaheuristics, clustering, missing data, adaptive optimization, prior knowledge

References:

[1] Goel, S., Tushir, M., Arora, J., Sharma, T., Gupta, D., Nauman, A., & Muhammad, G. (2024). *An Enhanced Integrated Method for Healthcare Data Classification with Incompleteness.* Computers, Materials & Continua, 81(2).

[2] Singh, J., & Gosain, A. (2024). *Revolutionizing Missing Data Handling with RFKFCM: Random Forest based Kernelized Fuzzy C-Means.* Procedia Computer Science, 33, 66-76.

[3] Saini, P., & Nagpal, B. (2024). *Analysis of missing data and comparing the accuracy of imputation methods using wheat crop data.* Multimedia Tools and Applications,83(14), 40393-40414.

[4] Davidovic, T., Glišovic, N., & Raškovic, M. (2018, July). *Bee colony optimization for clustering incomplete data.* In The 7th International Conference on Optimization Problems and Their Applications. OPTA-2018.

[5] Capariño, E. T. (2024, February). *Improving kNN Classification Performance with Correlated Attribute Imputation Mechanism.* In *Proceedings of the 2024 9th International Conference on Intelligent Information Technology* (pp. 287-294).

Hybrid CPU–FPGA Architectures for Efficient Execution of AI Algorithms

Aleksandar Peulić, Kristina Vasić

Faculty of Science, University of Kragujevac, Radoja Domanovića 12, 34000 Kragujevac, Serbia, aleksandar.peulic@pmf.kg.ac.rs,kristina.vasic@pmf.kg.ac.rs

Deep learning, a subfield of machine learning, relies on neural network models that require significant computational resources, especially when used in real-time applications such as image classification, speech recognition and autonomous systems. To address these challenges, hybrid architectures that combine traditional central processing units (CPUs) and field programmable gate arrays (FPGAs) have become increasingly important in recent years. These heterogeneous systems utilise the flexibility of CPUs with the parallelism and low-latency processing power of FPGAs and offer a powerful solution for accelerating inference tasks in artificial intelligence (AI). In this work, we present an overview and implementation concept for an AI inference accelerator that uses a CPU-FPGA co-design approach. The system offloads computationally intensive layers of a Convolutional Neural Network (CNN) to the FPGA fabric, while control and sequential tasks are managed by the CPU. The design includes modules for convolution, activation functions (ReLU, Tanh) and max-pooling, all mapped to dedicated hardware logic using multiply and accumulate (MAC) units and pipelined processing. This architecture is implemented on the PYNQ-Z2 platform, which integrates a Xilinx Zynq SoC that combines ARM Cortex-A9 processors and programmable logic in a single chip. The convolutional module uses a pipelined architecture to ensure continuous data flow and minimal latency, and utilises the FPGA's DSP slices for efficient fixed-point arithmetic. The pooling operations are controlled by a series of conditionally activated logic units that dynamically adapt to different input parameters. By utilising the inherent parallelism of FPGAs and offloading heavy computations from the CPU, this co-design approach achieves significant speed-ups and improved energy efficiency compared to CPU-only execution. The integration of CPU and FPGA in AI processing pipelines not only increases throughput and latency, but also enables deterministic execution times, which are critical for embedded and real-time applications. Therefore, hybrid architectures represent a promising direction for the use of AI at the edge of the system, especially in systems with limited power and resources. The adaptability of CPU-FPGA hybrid architectures enables easy customisation and scalability according to the specific requirements of the target application. This flexibility is particularly beneficial for rapidly evolving AI models and algorithms, where hardware accelerators need to keep pace with changing computational requirements without the need for a complete redesign. In addition, by utilising the partial reconfiguration capabilities of modern FPGAs, certain modules can be dynamically updated or replaced during runtime, enabling efficient resource utilisation and a longer device lifetime. These features make CPU FPGA systems ideal for use in edge devices such as autonomous drones, smart cameras and industrial IoT sensors, where power consumption, size and real-time performance are critical. As AI workloads continue to diversify, such heterogeneous computing platforms will play a key role in bridging the gap between high performance and energy efficiency in practical AI applications.

Keywords: hardware acceleration, hybrid architecture, convolutional neural networks, FPGA, CPU-FPGA co-design

References:

[1] Y. Qiao, J. Shen, T. Xiao, Q. Yang, M. Wen, and C. Zhang, “FPGA-accelerated deep convolutional neural networks for high throughput and energy efficiency,” *Concurrency and Computation: Practice and Experience*, vol. 29, no. 20, p. e3850, 2017

[2] Molina, R. S., Gil-Costa, V., Crespo, M. L., & Ramponi, G. (2022). High-level synthesis hardware design for FPGA-based accelerators: Models, methodologies, and frameworks. *IEEE Access, 10*, 90236–90254.

[3] Talib, M. A., Majzoub, S., Nasir, Q., Talpur, S., & Al-Rakhami, M. (2021). A systematic literature review on hardware implementation of artificial intelligence algorithms. *Journal of Supercomputing, 77*, 1897–1938.

[4] Trabelsi Ajili, M., & Hara-Azumi, Y. (2022). Multimodal neural network acceleration on a hybrid CPU–FPGA architecture: A case study. *IEEE Access, 10*, 9603–9617.
[5] Dhilleswararao, P., Boppu, S., Manikandan, M. S., & Cenkeramaddi, L. R. (2022). Efficient hardware architectures for accelerating deep neural networks: Survey. *IEEE Access, 10*, 131788–131828.

ASPEN: Solving Nonlinear Equality-Constrained Learning Problems in Machine Learning

Nemanja Vučićević[1], Nataša Krejić[2], Nataša Krklec Jerinkić[2], Tijana Ostojić[3]

[1] Faculty of Science, University of Kragujevac, Kragujevac, Serbia nemanja.vucicevic@pmf.kg.ac.rs,
[2] Faculty of Science, University of Novi Sad, Serbia natasak@uns.ac.rs
[2] Faculty of Science, University of Novi Sad, Serbia natasa.krklec@dmi.uns.ac.rs
[3] Faculty of Technical Sciences, University of Novi Sad, Serbia tijana.ostojic@uns.ac.rs

Motivated by the need for large-scale applications in machine learning, we propose ASPEN - a novel algorithm for solving potentially non-convex finite-sum optimization problems with nonlinear equality constraints, ASPEN belongs to a family of penalty-based methods which avoid potentially costly projections onto the feasible set. Moreover, in order to further reduce the computational cost, ASPEN uses an adaptive sampling strategy to determine the sample size dynamics. This strategy is based on an additional sampling approach proposed in [3] and further developed in several papers, including [2]. The IPAS algorithm [2] which covers only problems with linear equality constraints, while ASPEN covers also problems with nonlinear constraints and uses a different basic idea. We show the almost sure convergence of ASPEN method towards a stationary point under some standard assumptions for this framework.

To demonstrate the effectiveness of the method in real-world settings, we apply ASPEN to a constrained binary classification problem and compare it with two baseline strategies: a full sample method and a heuristic approach. We tested the algorithm on problems from the CUTEst library (Constrained and Unconstrained Testing Environment for Solvers), a widely used collection for evaluating optimization methods. To simulate a stochastic environment and adapt these deterministic problems to our algorithm, we introduce Gaussian noise into the original objective functions, creating perturbed versions of the problems. Experiments were conducted on these benchmark problems, with performance measured by computational efficiency metrics that reflect the overall cost of function evaluations. Results show that ASPEN achieves competitive or superior performance with significantly reduced computational effort, making it a promising tool for constrained optimization tasks in machine learning.

Keywords: Nonlinear Constrained Optimization; Machine Learning; Additional Sampling; Penalty method.

References:

[1] Gould, N. I., Orban, D., & Toint, P. L. (2015). CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Computational optimization and applications*, *60*, 545-557.
[2] Krejić, N., Jerinkić, N. K., Rapajić, S., & Rutešić, L. (2025). IPAS: An Adaptive Sample Size Method for Weighted Finite Sum Problems with Linear Equality Constraints. *arXiv preprint arXiv:2504.19629*.
[3] D. Di Serafino, N. Krejić, N. Krklec Jerinkić, M. Viola, LSOS: Line Search Second Order Stochastic Optimization Methods for Nonconvex Finite Sums, Mathematics of Computation, (2022).

Energy of Interval-Valued Bipolar Neutrosophic Soft Sets: A New Tool for Medical Decision-Making

Nemanja Vučićević[1], Nenad Stojanović[1], Marina Svičević[1], Filip Andrić[1]

Faculty of Science, University of Kragujevac, Kragujevac, Serbia
nemanja.vucicevic@pmf.kg.ac.rs, nenad.stojanovic@pmf.kg.ac.rs,
marina.svicevic@pmf.kg.ac.rs, filip.andric@pmf.kg.ac.rs

Many applications of neutrosophic soft sets have been limited by insufficient parametric tools (Dalkılıç, 2021). However, the hybridization of interval-valued bipolar neutrosophic sets (Deli, 2017) with soft sets has overcome this limitation, resulting in interval-valued bipolar neutrosophic soft sets — a structure with enhanced parametric richness and wide-ranging applicability. These sets have attracted significant attention for both theoretical investigation and practical problem-solving, particularly in decision-making contexts. To formalize this notion, consider a universal set U, and denote by $IVBNS(U)$ the collection of all interval-valued bipolar neutrosophic sets over U. Let E be a set of parameters (attributes) used to describe the elements of U, and let $A \subseteq E$ be a chosen subset of parameters relevant to the problem at hand.

An **interval-valued bipolar neutrosophic soft set (IVBNSS)** over U is then an ordered pair $(\mathcal{B}, A)$, where $\mathcal{B}$ is a mapping from E to $IVBNS(U)$. For each parameter $e \in E$, the mapping $\mathcal{B}(e)$ associates a set of ordered tuples of the form:

$$\left(u, \left[T^{+}_{L(u)}, T^{+}_{R(u)}\right], \left[I^{+}_{L(u)}, I^{+}_{R(u)}\right], \left[F^{+}_{L(u)}, F^{+}_{R(u)}\right], \left[T^{-}_{L(u)}, T^{-}_{R(u)}\right], \left[I^{-}_{L(u)}, I^{-}_{R(u)}\right], \left[F^{-}_{L(u)}, F^{-}_{R(u)}\right]\right),$$

for all $u \in U$. Here, the functions $T^{+}_{L}, T^{+}_{R}, I^{+}_{L}, I^{+}_{R}, F^{+}_{L}, F^{+}_{R} : U \to [0,1]$, represent the lower and upper bounds of the degrees of truth, indeterminacy, and falsity in the positive domain, while $T^{-}_{L}, T^{-}_{R}, I^{-}_{L}, I^{-}_{R}, F^{-}_{L}, F^{-}_{R} : U \to [-1,0]$, capture the corresponding degrees in the negative domain.

In the work of Mudrić Staniškovski et al., the authors introduced an energy measure for fuzzy soft sets by constructing real-valued matrices based on membership degrees and computing their nuclear norm — the sum of singular values. This measure effectively captured the structural contribution of individual elements to the soft set system and was successfully applied in decision-making scenarios involving uncertainty. The method emphasized the spectral structure of the system rather than relying solely on aggregate membership scores, thereby providing more nuanced insights.

Building on this spectral perspective, we extend the energy concept to the interval-valued bipolar neutrosophic soft setting by incorporating both positive and negative information, represented through real-valued matrices constructed from the lower and upper bounds of the truth, indeterminacy, and falsity functions. The energy of an interval-valued bipolar neutrosophic soft set (IVBFNSS) is computed based on the singular values of matrices formed from the lower and upper bounds of the positive and negative truth, indeterminacy, and falsity functions. For each alternative, separate real-valued matrices are constructed for each component, including indeterminacy and falsity. Subsequently, we compute the optimistic and pessimistic energy values for each alternative (both based on combinations of the lower and upper bounds of all functions). The resulting energy is defined as the arithmetic mean of these two values. By performing these computations, we derive a ranking of the alternatives based on their energy values, allowing us to determine the relative ordering and identify the optimal solution.

The performance of the proposed algorithm is evaluated with a particular focus on medical diagnosis problems — a domain where precise and reliable decision-making is critically important. This is especially relevant in frequently studied scenarios, such as the assessment of individuals based on

parameters like happiness, sadness, energy levels, and sleep rate, used to identify conditions such as bipolar disorder.

To briefly highlight the details of the proposed approach and the comparative analysis: the energy-based method within the interval-valued bipolar neutrosophic soft set framework demonstrates notable advantages, particularly in decision-making contexts. A comparative analysis with existing algorithms — especially the method by Abdel-Basset et al. (2019), a key contribution in the field — highlights the superior efficiency and robustness of our approach. The algorithm was applied to several medical diagnosis examples to determine the best alternative, where it performed successfully. Unlike the Abdel-Basset et al. algorithm, which failed to identify the best alternative in a specific case, our method consistently selected the optimal alternative and successfully determined the correct solution.

Moreover, the proposed algorithm shows strong potential for handling large-scale, high-dimensional medical data.

Keywords: Interval-valued bipolar neutrosophic soft set, Decision making, Singular value, Nuclear norm

References:

[1] Abdel-Basset, M., Mohamed, M., Elhoseny, M., Son, L. H., Chiclana, F., & Zaied, A. E. N. H. (2019). Cosine similarity measures of bipolar neutrosophic set for diagnosis of bipolar disorder diseases. *Artificial intelligence in medicine*, *101*, 101735.

[2] Dalkılıç, O. (2021). Relations on neutrosophic soft set and their application in decision making. *Journal of Applied Mathematics and Computing*, 67, 257–273.

[3] Deli, I. (2017). Interval-valued neutrosophic soft sets and its decision making. *International Journal of Machine Learning and Cybernetics*, *8*, 665-676.

[4] Mudrić Staniškovski, Lj., Djurović, Lj., & Stojanović, N. (2024). Energy of a fuzzy soft set and its application in decision-making. *Iranian Journal of Fuzzy Systems*, 21(2), 35–49.

AI-Powered Recommender System for Opioid Recovery: A Personalized Approach to Behavioral Health Intervention

Tanuja Voruganti[1], Faizan Raza[1], Mark Fahim[1], Nonna Sorokina[1], Dusan Ramljak[1], Tatjana Davidovic[2], Dragutin Ostojic[3]

[1]School of Professional Graduate Studies at Great Valley, The Pennsylvania State University, Malvern, PA, USA, dusan@psu.edu
[2]Mathematical Institute, Serbian Academy of Sciences and Arts, Belgrade, Serbia, tanjad@mi.sanu.ac.rs
[3]Faculty of Science, University of Kragujevac, Kragujevac, Serbia, dragutin.ostojic@pmf.kg.ac.rs

The opioid crisis remains a major public health challenge in the United States, contributing to over 80,000 overdose deaths annually (Ahmad et al., 2025). Despite growing access to treatment, many individuals in recovery lack personalized and timely support. This paper presents an AI-powered Minimum Viable Product (MVP) developed during the Pennsylvania State University Nittany AI Challenge (https://nittanyai.psu.edu/alliance-programs/nittany-ai-challenge/), offering a recommender system that provides tailored behavioral health strategies based on user input. The goal is to empower early intervention through scalable, data-driven, and user-centric digital tools. MVP is designed to meet individuals where they are—offering not just information, but understanding, support, and personalized guidance. At its heart, this MVP is more than just a web application. It's a small digital ally that walks alongside someone who may be struggling, helping them better understand their relationship with opioids and take manageable steps toward harm reduction and self-care.

Machine learning models such as neural networks have shown considerable promise in capturing non-linear patterns in user behavior and health interventions (He et al., 2017). This system builds on those findings, using cosine similarity for profile matching. We employed this methodology over Euclidean distance or graph-based methods (e.g., GNNs) for three main reasons. First is for directional alignment, since cosine similarity measures profile similarity regardless of response magnitude (e.g., users scoring 1-3 vs. 3-5 on Likert scales), which is crucial for comparing addiction severity patterns. Secondly, cosine similarity is computationally lightweight, making it suitable for MVP deployment, whereas graph-based methods, GNNs in particular, require complex graph infrastructures. Finally, cosine similarity is a critical feature for behavioral health data (Chen et al., 2020).

This project demonstrates how artificial intelligence can be ethically applied to sensitive behavioral health domains using surrogate data, offering both feasibility and promise. The MVP lays a foundation for digital recovery tools capable of empowering person-centered intervention strategies across substance use disorders (Volkow, Jones, Einstein, & Wargo, 2019).

User engagement is driven through a four-step web interface developed using Flask: a landing page introducing the opioid crisis, a behavioral quiz collecting user input (see Figure 1), a recommendation page with tailored interventions, and a feedback module to assess usefulness. The cosine similarity algorithm compares users to the most behaviorally similar profiles and recommends corresponding tips. For example, users classified as *high severity* should receive evidence-based suggestions such as "Consult a healthcare provider," while *moderate severity* users were offered self-guided strategies like "Practice mindfulness" or "Establish a support routine."

The proposed system integrates a Multi-Layer Perceptron (MLP) classifier with a cosine similarity engine to assess risk severity and generate customized recovery tips. The MLP classifier, configured with hidden layers of (16, 8) neurons and the ReLU activation function, trained on a synthetic dataset

simulating opioid usage patterns and behavioral traits (e.g., social support, impulsivity) of 100 user profiles, achieved 100% training and 91% testing accuracy using a 80/20 train-test split. The similarity scores for the top five nearest neighbors ranged from 0.75 to 0.95, indicating a high degree of relevance in the recommendations. These synthetic datasets, modeled from validated psychological scales, offer an ethically sound alternative when real opioid usage data is restricted due to privacy regulations (Elmagarmid et al., 2014).

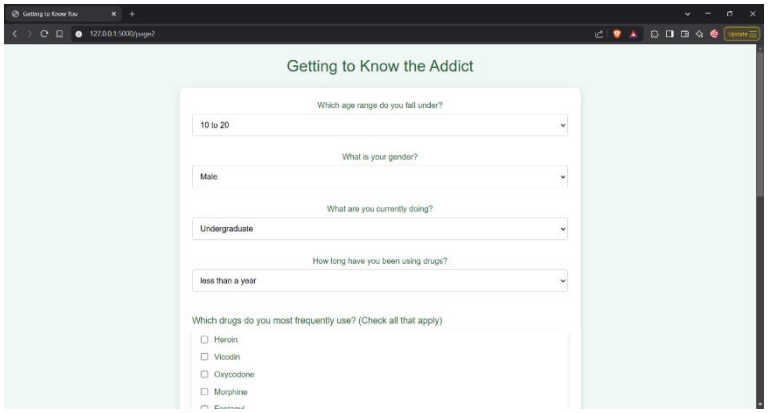


Fig 1 Behavioral quiz page for gathering user info

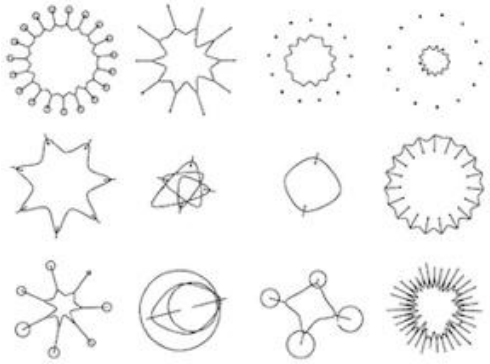
This AI-powered recommender system MVP demonstrates a proof-of-concept AI system designed to address the multifaceted opioid crisis by gathering user data on opioid usage behaviors, classifying the severity of a user's condition, and recommending personalized tips or strategies. Built with a clear and user-friendly four-page flow, the system leverages an internal architecture featuring an MLP model and a similarity engine to lay the groundwork for tailored recommendations. This AI-powered recommender system highlights a transformative approach to opioid recovery support, harnessing user interaction data and similarity-based insights to predict and personalize rehabilitation pathways. With more data, refinements, and user feedback, this solution can evolve into a powerful, data-driven tool that not only empowers individuals on their recovery journey but also optimizes intervention strategies for healthcare providers. The MVP serves as a scalable foundation for combating opioid addiction, supporting long-term recovery efforts, and potentially curbing excessive social media usage as an additional application. The system was tested in a local deployment environment, where feedback from simulated users affirmed the interface's usability and the recommendations' relevance. Limitations include the static nature of the tips and the synthetic training data, but the feedback module provides a mechanism for future adaptability. Future work will integrate real clinical datasets, enable dynamic recommendation learning, and address ethical deployment considerations—particularly around privacy and data interpretability.

Keywords: Machine learning, neural networks, recommender systems, behavioral health, opioid recovery

References:

[1] Ahmad FB, Cisewski JA, Rossen LM, Sutton P. (2025). Provisional drug overdose death counts. National Center for Health Statistics. DOI: https://dx.doi.org/10.15620/cdc/20250305008
[2] Chen, J., Wang, Y., Huang, J., & Wu, Z. (2020). Personalized recommendation in health care using machine learning techniques. IEEE Access, 8, 151164-151176.
[3] Elmagarmid, A. K., et al. (2014). Synthetic data in healthcare. Journal of the American Medical Informatics Association, 21(1), 1-9.
[4] He, X., Liao, L., Zhang, H., et al. (2017). Neural collaborative filtering. Proceedings of the 26th International Conference on WWW, 173-182.
[5] Volkow, N. D., Jones, E. B., Einstein, E. B., & Wargo, E. M. (2019). Prevention and treatment of opioid misuse and addiction: A review. JAMA Psychiatry, 76(2), 208-216.

EXPLAINABLE ARTIFICIAL INTELLIGENCE



The incorporation of LLM tool-calling capabilities for enhancing the Explainable AI paradigm in classification tasks

Mina Nikolić[1], Aleksandar Stanimirović[2] and Leonid Stoimenov[3]

[1]Department of Computer science and informatics, Faculty of Electronic Engineering, University of Niš, Serbia, mina.nikolic@elfak.ni.ac.rs
[2]Department of Computer science and informatics, Faculty of Electronic Engineering, University of Niš, Serbia, aleksandar.stanimirovic@elfak.ni.ac.rs
[3]Department of Computer science and informatics, Faculty of Electronic Engineering, University of Niš, Serbia, leonid.stoimenov@elfak.ni.ac.rs

The use of Large language models in recent times has substantially transformed the landscape of Natural language processing and led to visible innovation in various fields such as medicine, education, law, research, engineering and others.

Common tasks such as text generation, summarization, question answering and language translation are now considered to be only a subset of the capabilities that Large language models can exhibit. With modern research trends such as the paradigm shift from Generative Artificial Intelligence to Agentic Artificial Intelligence, the focus is directed towards incorporating execution capabilities in various degrees of autonomy (Schneider, 2025).

The potential of their incorporation is already visible in various scenarios regarding API calls meant for external usage, such as search engines and custom function execution. Knowing that, the possibility for enabling the Explainable AI paradigm in these kinds of workflows seems like a plausible solution for tackling the black-box nature of ML model predictions.

The main motivation behind this research lies in the sense that Large language models can be used to mitigate the barrier between the explanations of decisions made by ML models and clear natural language explanations generated by the language models itself.

The kind of explainability possible when considering Large language models can be viewed from various points, including the analysis of the language model itself (Fujiwara et al., 2024), the language model acting as an explainer of a specific model (Kroeger et al., 2023) or as a tool for enhancing the explainability provided by the existing XAI mechanism (Zytek et al., 2024).

The concrete research problem analyzed is regarding the multi-class classification of student performance in the education domain, while utilizing predictions gathered by LIME (Ribeiro et al., 2016) as the backbone for enabling even further explainability with the textual based language model explanations. By doing so, another layer of explanations can be added, eliminating the need for understanding the underlying mechanism of XAI.

The steps needed to ensure the proper functioning of such a system are defined as follows: dataset analysis and preprocessing, training and evaluation of multiple candidate ML models, defining functions for local, instance-based explainability, defining function calling tools for the Large Language model and adequately prompting the Large language model to ensure proper explanations.

The dataset contains information regarding the factors of academic success in high school students (El Kharoua, 2024). The input variables reflect the demographic details, study habits, parental involvement and extracurricular activities, while the target variable is set to be the grade point average (GPA) split into 5 categories. For the purpose of multi-class classification, the best performance was obtained with the Random Forest classifier, while the Large language model of choice was OpenAI's GPT-4.1.

The initial results gathered from the proposed explainability pipeline are intended to be analyzed in terms of a single instance at a time, which is influenced by the choice of utilizing local explainability. The results for a particular test instance are displayed in Figure 1.

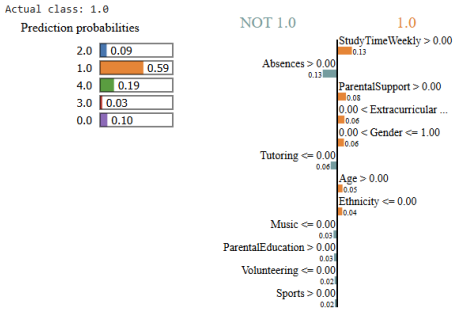


Fig. 1: Explainability of a test instance using LIME

The actual class of the instance and the prediction made by LIME's surrogate model are in alignment, while also having meaningful positive and negative attribute influences towards the predicted class. That is not always the case, as the surrogate model for approximation tends to be far simpler than the actual prediction model. The Large language model enhanced explainability for the same instance is displayed in Figure 2.

Here's how the values you provided influenced the model's classification decision:

Largest positive influences (pushed towards the predicted class):

1. **StudyTimeWeekly > 0.00** (+0.1279) – Spending time studying weekly contributed the most in favor of this classification.
2. **ParentalSupport > 0.00** (+0.0763) – Parental support also played a significant positive role.
3. **0.00 < Extracurricular ≤ 1.00** (+0.0630) – Participation in extracurricular activities added some positive weight.
4. **0.00 < Gender ≤ 1.00** (+0.0589) – The gender attribute provided some positive influence.
5. **Age > 0.00** (+0.0464) – The age attribute contributed positively.
6. **Ethnicity ≤ 0.00** (+0.0437) – This particular category of ethnicity had a positive effect as well.

Largest negative influences (pushed against the predicted class):

1. **Absences > 0.00** (−0.1263) – Having absences was the strongest negative influence, pushing against the predicted class.
2. **Tutoring ≤ 0.00** (−0.0569) – Not participating in tutoring was another notable negative factor.
3. **Music ≤ 0.00** (−0.0325) – Not taking part in music activities had a mild negative effect.
4. **ParentalEducation > 0.00** (−0.0282) – Having parents with higher educational backgrounds slightly pushed against the classified class.
5. **Volunteering ≤ 0.00** (−0.0226) – Not volunteering had a small negative impact.
6. **Sports > 0.00** (−0.0207) – Participating in sports slightly opposed the classification, but this effect was minor.

Summary:
The decision was primarily driven by strong positive influences from study time, parental support, and extracurricular participation, which outweighed the significant negative impact from having absences. The remaining features had smaller effects in either direction.

If you want a specific explanation for the class label itself, let me know!

Fig. 2: Explainability of a test instance enhanced by the LLM

The explanations provided by the Large language model are directly influenced by the predictions gathered from LIME and serve as an added layer of explainability. By doing so, a natural language explanation can be obtained, which is especially beneficial for understanding the logic behind the influence of specific attributes on the class of choice. The complete code for the implementation can be seen at [https://github.com/minanikolic916/XAI_LLMs].

Acknowledgments: This work has been supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia [Grant Number: 451-03-137/2025-03/ 200102].

Keywords: Explainable Artificial Intelligence (XAI), Large language models (LLMs), Local Interpretable Model-Agnostic Explanations (LIME), Machine Learning (ML), data classification

References:

[1] Schneider, J. (2025). Generative to agentic ai: Survey, conceptualization, and challenges. *arXiv preprint arXiv:2504.18875*.

[2] Fujiwara, K., Sasaki, M., Nakamura, A., & Watanabe, N. (2024). Measuring the interpretability and explainability of model decisions of five large language models. *Open Science Framework: Charlottesville, VA, USA*.

[3] Kroeger, N., Ley, D., Krishna, S., Agarwal, C., & Lakkaraju, H. (2023). Are large language models post hoc explainers?.

[4] Zytek, A., Pidò, S., & Veeramachaneni, K. (2024). Llms for xai: Future directions for explaining explanations. *arXiv preprint arXiv:2405.06064*.

[5] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
[6] El Kharoua, R.: Students Performance Dataset, https://doi.org/10.34740/KAGGLE/DS/5195702, 2024.

Explainable AI in Medicine: Bridging Trust, Transparency, and Clinical Utility

Aleksandra Stojnev Ilic[1], Dragan Stojanovic[1], Milos Ilic[2], Natalija Stojanovic[1]

[1]Faculty of Electronic Engineering, University of Nis, Serbia,
aleksandra.stojnev@elfak.ni.ac.rs dragan.stojanovic@elfak.ni.ac.rs
natalija.stojanovic@elfak.ni.ac.rs
[2]Toplica Academy of Applied Sciences, Prokuplje, Serbia, milos.ilic@vpskp.edu.rs

Artificial intelligence (AI) is increasingly integrated into modern healthcare systems, providing advanced tools for diagnosis, prognosis, and clinical decision support. Despite strong performance, many machine learning models, especially deep learning, lack transparency, which limits clinical trust, patient communication, and regulatory approval. Explainable AI (XAI) addresses these issues by making model predictions interpretable and trustworthy for both technical and non-technical users. Explainability is essential for ethical and safe deployment, enabling clinicians to understand the rationale behind predictions in critical decisions such as treatment planning and triage. Additionally, XAI supports error detection, bias auditing, and system debugging in complex real-world settings. The development of XAI in medicine draws on computer science for algorithm design, interpretability frameworks, and secure infrastructure, alongside machine learning techniques for transparent modeling. Effective adoption also requires integration with clinical workflows via robust software engineering and user-centered design, resulting in AI systems that are accurate, transparent, and practical for routine clinical use.

XAI methods in healthcare can be broadly categorized into post-hoc explanation techniques and intrinsically interpretable models. Post-hoc methods operate on already-trained models and include feature attribution techniques such as SHAP and LIME, which estimate the contribution of input variables to a prediction. In medical imaging, gradient-based saliency methods (e.g., Grad-CAM) are commonly used to highlight regions in radiographs or pathology slides that influence decisions. Counterfactual explanations have also gained traction, allowing clinicians to understand how minor changes in input (such as lab results or symptom profiles) would alter an AI-driven recommendation. In contrast, intrinsically interpretable models are designed to be transparent from the outset. Examples include decision trees, rule-based systems, and generalized additive models (GAMs). While these models are more understandable, they often underperform compared to black-box alternatives on complex, high-dimensional medical data. Hybrid approaches attempt to balance these trade-offs by incorporating interpretable structures (like attention mechanisms or case-based reasoning) into otherwise complex models.

XAI has been successfully deployed in several areas of medicine. In radiology, pathology, and dermatology, visual explanations help verify whether models are focusing on diagnostically relevant regions, exposing cases of “shortcut learning” based on spurious features. In critical care, interpretable early warning systems help prioritize patients based on transparent risk factors. Personalized medicine also benefits, with explainable models clarifying the drivers of individual treatment recommendations. In mobile health, XAI improves patient engagement by making insights from wearable data more transparent and actionable. Building on this foundation, our previous work demonstrated how XAI techniques, specifically **SHAP and LIME,** can be applied to continuous glucose monitoring (CGM) data to enhance both the interpretability and trustworthiness of predictive models. We use SHAP to quantify the contribution of recent glycemic patterns and contextual features to predicted glucose anomalies, and LIME to generate localized, instance-specific explanations for individual prediction outputs. These explanations allow users and clinicians to understand which features most influenced each anomaly detection, facilitating insight into unlogged lifestyle events such as exercise, fasting, or hormonal changes. If combined with unsupervised techniques like UMAP and DBSCAN for anomaly visualization, our approach can offer a comprehensive framework for explainable, personalized, and clinically relevant CGM analysis.

Despite progress, several challenges limit the widespread adoption of XAI in medicine. First, there is a well-known trade-off between model accuracy and interpretability. Complex models tend to outperform simpler ones but at the cost of opacity. Second, explanations must be clinically relevant; technical fidelity alone is insufficient if the outputs are unintelligible or misaligned with clinical reasoning. Third, user variability presents a barrier: an explanation suitable for a data scientist may not meet the needs of a general practitioner or a patient. Moreover, explanation methods themselves can introduce biases or create false reassurance, particularly when they are misinterpreted as evidence of causality. Finally, regulatory bodies have yet to standardize what constitutes a valid or sufficient explanation, leading to uncertainty among developers and healthcare providers alike. Future research must prioritize human-centered and context-aware explainability strategies. One promising direction involves incorporating causal reasoning into XAI to move beyond correlation-based insights. Another involves the development of federated and privacy-preserving explanation methods that respect the sensitive nature of medical data while still enabling transparency across distributed systems. Interactive explanation interfaces, which allow clinicians to query model reasoning or simulate counterfactual scenarios, also show promise in improving usability. Finally, the field would benefit from standardized evaluation metrics that go beyond technical validity to assess explanation quality, trust calibration, and clinical impact.

Explainable AI stands at the intersection of data science, clinical medicine, ethics, and regulation. As healthcare systems increasingly turn to AI for support, the need for transparent, interpretable, and clinically meaningful systems becomes critical. XAI offers a path forward by enabling trust, improving safety, and supporting informed decision-making. However, achieving explainability in practice will require interdisciplinary collaboration, methodological innovation, and close attention to real-world clinical workflows. The future of AI in medicine will not be defined by accuracy alone, but by our ability to understand and act upon the predictions these systems provide.

Acknowledgement: This work has been supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia [Grant Number: 451-03-137/2025-03/ 200102]

Keywords: Explainable Artificial Intelligence (XAI), Medical AI, Clinical Decision Support, Model Interpretability

References:

[1] Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine, Nature Medicine, vol. 28, pp. 31-38.
[2] Tjoa, E., & Guan, C. (2020). A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI, IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4793-4813.
[3] Schoonderwoerd, T. A.J., Jorritsma, W., Neerincx, M. A., & Bosch, K. (2021). Human-centered XAI: Developing design patterns for explanations of clinical decision support systems, International Journal of Human-Computer Studies, vol. 154:102684.
[4] Sun, Q., Akman, A., Schuller, B. W. (2025). Explainable Artificial Intelligence for Medical Applications: A Review, ACM Transactions on Computing for Healthcare, vol. 6, Issue 2, Article No.: 17, pp. 1 – 31.
[5] Stojnev Ilic, A., Stojanovic, D., Stojanovic, N., Ilic, M. (2025). Explainable AI for Continuous Glucose Monitoring Data, 4th Serbian International Conference on Applied Artificial Intelligence (SICAAI).

Differences between explainability in tabular and unstructured data: An experimental analysis of XAI methods

Enisa Trubljanin[1,2], Nataša Kovač[2]

[1]Faculty for Information Systems and Technologies, University of Donja Gorica, udg@udg.edu.me
[2]Faculty of Applied Sciences, University of Donja Gorica, udg@udg.edu.me

In modern artificial intelligence systems, model explainability has become crucial for increasing user trust, identifying biases, and supporting reliable decision-making, especially in critical domains such as medicine and security. This paper explores the differences in explainability approaches between tabular and unstructured data through a comparative application of contemporary XAI techniques. For tabular data, an XGBoost model was used along with SHAP analysis, while for unstructured (visual) data, a convolutional neural network was applied using Grad-CAM visualizations. The results show that local explanations in tabular data are more direct and interpretable, whereas explanations in image data require additional context and interpretation. The advantages and limitations of both approaches are discussed, along with practical recommendations for selecting XAI methods based on data structure and target user group. The paper emphasizes the need for domain-specific and multi-layered explainability strategies in the development of transparent AI systems.

Keywords: Explainability, XAI techniques, Tabular and unstructured data, Transparent AI

References:

[1] Sahakyan, M., Aung, Z., & Rahwan, T. (2021). Explainable artificial intelligence for tabular data: A survey. *IEEE access*, *9*, 135392-135422.
[2] Vilone, G., & Longo, L. (2021). Classification of explainable artificial intelligence methods through their output formats. *Machine Learning and Knowledge Extraction*, *3*(3), 615-661.
[3] O'Brien Quinn, H., Sedky, M., Francis, J., & Streeton, M. (2024). Literature review of explainable tabular data analysis. *Electronics*, *13*(19), 3806.
[4] Gupta, J., & Seeja, K. R. (2024). A comparative study and systematic analysis of XAI models and their applications in healthcare. *Archives of Computational Methods in Engineering*, *31*(7), 3977-4002.
[5] Hossain, M. I., Zamzmi, G., Mouton, P. R., Salekin, M. S., Sun, Y., & Goldgof, D. (2025). Explainable AI for medical data: current methods, limitations, and future directions. *ACM Computing Surveys*, *57*(6), 1-46.

Explainable Autonomous Driving Research with Object-Level Vector Modality Integration

Jinyuan Zhang[1], Lei Zhang[2], Jianxun Cui[3*]

[1]College of Transportation, Jilin University, China, jyzhang1721@mails.jlu.edu.cn
[2]School of Transportation Science and Engineering, Harbin Institute of Technology, China, 2022110963@stu.hit.edu.cn
[3]School of Transportation Science and Engineering, Harbin Institute of Technology, China, cuijianxun@hit.edu.cn

This paper presents an object - level vector multimodal LLM architecture to boost the interpretability of autonomous driving decisions and strengthen semantic understanding and reasoning of driving scenarios. A large - scale driving dataset with 10k scenes and 160k Q & A pairs was built via reinforcement learning and GPT - 3.5 labeling, achieving cross - modal alignment and efficient supervision. Using a two - stage training strategy, the model first maps numerical modalities to the semantic space and then fine - tunes on driving Q & A to enhance action prediction and explanation. Experiments show it outperforms traditional behavioral cloning methods, proving LLMs' potential in autonomous driving.

Introduction: Large language models (LLMs) offer a new paradigm for transparent autonomous driving decision - making. This paper presents an object - level vector multimodal LLM architecture. It uses structured vectors to represent key driving - scene elements, enabling efficient semantic modeling and interpretable - decision – making (Wang et al., 2023). A 10k - driving - scene dataset was built to generate 160k Q & A pairs, and a two - stage training strategy was adopted. The architecture's innovations lie in its cost - effective computation, enhanced decision traceability, and being the first to create a driving - specific Q & A scoring system. Experiments show it surpasses conventional methods in computational efficiency and interpretability.

Methodology: Reinforcement Learning Data Collection: A 2D simulator was built to create high - fidelity scenes with roads, vehicles, and pedestrians. Object - level vector representations of key information were designed. A reinforcement learning agent with the PPO algorithm collected data from 15 traffic conditions, forming a cross - modal dataset. Structured Language Transformation: A Trans function was designed to convert object - level vectors into pseudo - linguistic labels (Li et al., 2023). Two versions of labels were generated for pre - training and fine - tuning, guiding the model to link driving decisions with semantic explanations. Automated Question - Answer Data Collection: The Trans function converted scene vectors into natural - language descriptions. Combined with GPT, it generated question - answer pairs covering six types of questions, with safeguard rules for edge cases. Model Training: A two - stage training strategy was proposed. Driving - scene vectors were incorporated into LLaMA - 7B to build an interpretable driving - decision system, achieving decision - making transparency.

Evaluation:

1.Perception and Action Prediction Evaluation

Based on a 1k-scenario evaluation set, we compared three models: two-phase trained Driver-LLM; fine-tuning-only Driver-LLM; Perceiver model (with equivalent parameter size, replacing the LLM with a policy module). Evaluation metrics included perception accuracy (vehicle/pedestrian count errors, traffic light detection accuracy, and distance error), action prediction performance (longitudinal/lateral control errors), and semantic generation quality (token loss).

Table 1. Evaluation Results of Perception and Action Prediction

	E_{car}	E_{ped}	Acc_{TL}	D_{TL}	E_{lon}	E_{lat}	L_{token}
Driver-LLM w/o pretrain	**0.0657**	**0.3131**	0.7179	6.6241	**0.0666**	0.0144	**0.5015**
Driver-LLM w/ pretrain	0.1010	1.6677	0.7576	7.4751	0.0939	0.0144	0.6437
Perceiver	0.8693	0.6841	**0.9000**	**0.4099**	0.1797	0.1115	n/a

Results indicate that pretraining with semantic alignment significantly enhances the LLM's scene understanding and decision interpretability, whereas the Perceiver, lacking language reasoning, underperformed in tasks requiring commonsense judgment.

2.Driving QA Quality Assessment

We employed two scoring mechanisms: GPT-3.5-based automated scoring (0–10 points) by inputting scene descriptions, questions, and model responses; human validation via manual scoring of 100 randomly sampled QA pairs to mitigate GPT's bias toward semantically similar answers (Bubeck et al., 2023). Results demonstrated that pretrained Driver-LLM significantly outperformed non-pretrained models in both GPT and human scores, proving that pretraining effectively strengthens the model's understanding of driving semantics and natural language explanations.

Table 2. Grading of Driving QA Outputs

	GPT Grading	Human Grading
Driver-LLM w/ pretrain	8.38	7.58
Driver-LLM w/o pretrain	7.51	6.83

Conclusion: The research limitations are as follows. First, the model is based on open - loop static scenarios, failing to adequately simulate dynamic environmental changes. Second, existing vector representations can't fully capture complex road topology and dynamic trajectory predictions. Third, GPT scoring is overly lenient towards logically flawed answers, and there's limited sampling for manual verification. In the future, the model architecture will be optimized, high - order vector representations designed, and a multi - dimensional quantitative evaluation system established. LLMs are of key value in enhancing the interpretability of autonomous driving systems.

Keywords: Deep Learning, Autonomous Driving, Large Language Models (LLMs), Interpretability, Driving Scenarios

References:

[1] Bubeck, S., Chadrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023, March). Sparks of artificial general intelligence: Early experiments with gpt-4.

[2] Kim, J., Moon, S., Rohrbach, A., Darrell, T., & Canny, J. (2020). Advisable learning for self-driving vehicles by internalizing observation-to-action rules. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9661-9670).

[3] Wang, X., Chen, G., Qian, G., Gao, P., Wei, X. Y., Wang, Y., ... & Gao, W. (2023). Large-scale multi-modal pre-trained models: A comprehensive survey. Machine Intelligence Research, 20(4), 447-482.

GENERAL AND GENERATIVE ARTIFICIAL INTELLIGENCE

Risk, Regulation, and Readiness: Generative AI Adoption in Healthcare and IT

Aleksandra Klašnja-Milićević[1], Mirjana Ivanović[1], Jelena Ilić[1], Nataša Papić-Blagojević[2]

[1]Faculty of Sciences, University of Novi Sad, Serbia, akm@dmi.uns.ac.rs, mira@dmi.uns.ac.rs, ilicjelena3105@gmail.com
[2]Novi Sad School of Business, Serbia, npapic.blagojevic@gmail.com

Generative Artificial Intelligence (GAI) tools have rapidly emerged as transformative technologies across professional domains. This paper presents a comparative analysis of GAI deployment in two contrasting domains: healthcare and information technology (IT). Drawing on survey data from 173 professionals and recent scholarly literature (2023–2025), we examine how risk perception, regulatory environment, and organisational readiness shape GAI adoption in each sector. The survey instrument, designed to support descriptive and inferential analysis, combined multiple-choice, Likert-scale, and open-ended questions to capture quantitative and qualitative insights into adoption trends, ethical concerns, and institutional practices.The healthcare sample covered diverse specializations (e.g., neurology, cardiology), while the IT group involved software engineers, analysts, and system administrators.

In healthcare, adoption of GAI is progressing cautiously. Only 19.6% of surveyed professionals reported prior usage of tools such as ChatGPT in clinical or administrative workflows. GAI is primarily used for drafting medical reports (e.g., discharge summaries), simplifying clinical language for patient communication, and automating documentation. However, broader integration is hindered by stringent data protection laws (e.g., HIPAA, GDPR), ethical uncertainty, lack of clinical validation, and professional liability concerns. Physicians and administrators emphasize the need for explainable AI systems and institutional approval processes before use in decision-making. These findings align with Heller and Lee (2024), who report that the success of AI implementation in healthcare depends on comprehensive governance frameworks and stakeholder trust. Zhao et al. (2024) further caution that GAI tools introduce unique risks for patient privacy, particularly through inadvertent data exposure.

In IT, over 70% of respondents report frequent GAI use, primarily for code generation, debugging, and documentation. Commonly used tools include GitHub Copilot for code suggestions, Tabnine for IDE integration, and ChatGPT for logic review and documentation enhancement. While users cite speed and efficiency gains, they also express concerns about hallucinated outputs, AI-generated responses that appear coherent but are factually incorrect (e.g., plausible yet invalid API syntax), as well as the traceability of AI-generated code and over-reliance on such tools without proper validation. Our data show that developers who rely heavily on GAI tend to favour internal review protocols over external regulation. Russo (2023) identifies these challenges as part of a shifting paradigm in software engineering, where critical thinking and review mechanisms must evolve alongside AI capabilities. Unlike healthcare, IT professionals operate within more agile environments with fewer regulatory constraints, enabling faster adoption but demanding stronger internal governance.

Despite contextual differences, both sectors emphasize the need for clear guidelines and ongoing AI training. However, healthcare respondents report lower access to such resources, highlighting a readiness gap that may explain lower adoption. The study also revealed significant differences in perceived benefits. In healthcare, the value of GAI is framed around efficiency and patient communication support, with limited tolerance for error. In IT, benefits are seen in speed, flexibility, and ideation, with a higher risk tolerance. This divergence reflects broader epistemic and operational differences: healthcare prioritizes accuracy, traceability, and safety; IT privileges agility, performance, and creativity. These findings suggest that GAI adoption must be guided by sector-specific strategies. In healthcare, pathways for clinical validation, ethical oversight, and legal accountability are essential. In IT, quality assurance mechanisms, documentation standards, and human oversight should remain

integral. Policymakers and institutional leaders should invest in context-aware governance models and foster intersectoral learning. Furthermore, it is crucial to consider the role of domain-specific data quality and availability. Healthcare institutions often struggle with siloed, incomplete, or non-standardized datasets, which can undermine the performance of GAI models trained on clinical records. In contrast, IT organizations typically have greater control over data pipelines and the ability to synthetically generate large training corpora. Addressing these disparities through improved data infrastructure, common interoperability standards, and shared repositories can enhance the fairness and efficacy of GAI applications in both sectors. Preliminary analyses included cross-tabulations, χ^2 tests, and Spearman's correlations. In the healthcare group, familiarity with GAI systems showed a moderate positive correlation with comfort in using them ($\rho = 0.42$, $p < .01$). Furthermore, those who perceived GAI as effective in supporting patient communication were significantly more open to receiving formal training. In IT, prior experience with GAI tools (e.g., ChatGPT) strongly predicted both frequency of use and perceived utility ($\chi^2 = 12.6$, $p < .01$). These findings will be elaborated with visualizations and explored through exploratory cluster analysis to identify dominant user profiles in both sectors.

In the conference presentation, we will present quantitative results from the survey, including cross-tabulated comparisons between sectors. Specific findings include the disparity in adoption rates (19.6% in healthcare vs. over 70% in IT), differing perceptions of institutional readiness, and risk scoring (mean concern scores: 4.6 in healthcare vs. 3.1 in IT). Additionally, we are preparing exploratory correlation analyses between risk perception and self-reported adoption frequency, as well as thematic analysis of open-ended responses to identify patterns in ethical concerns. This study underscores the importance of tailoring Generative AI implementation strategies to the specific needs and constraints of individual sectors. While healthcare prioritizes accuracy, explainability, and institutional oversight, the IT sector benefits from a more experimental and agile approach. Despite these differences, both domains express a shared demand for ethical guidance, user training, and robust evaluation frameworks. The stark contrast in adoption rates and institutional readiness highlights the need for nuanced governance models that reflect domain-specific priorities.

Future research should expand on these findings by incorporating longitudinal data to track changes in adoption patterns over time. Additionally, sector-specific case studies could provide deeper insights into the organizational, technical, and cultural factors that influence successful integration of GAI tools. Comparative analysis across additional sectors, such as education, law, or finance, may further enrich our understanding of the conditions under which GAI can be adopted responsibly and effectively. Finally, integrating perspectives from patients, clients, or end-users can offer a more holistic view of the societal impact of generative technologies.

Keywords: Generative AI, Healthcare, Information Technology, AI Governance, Technology Adoption

References:

[1] Heller, A., & Lee, K. (2024). Adoption of artificial intelligence in healthcare: A survey of global health systems. Journal of the American Medical Informatics Association, 31(3), 455–463. https://doi.org/10.1093/jamia/ocaf065

[2] Zhao, Y., Smith, R., & McKenna, B. (2024). Generative AI in medical practice: An in-depth exploration of privacy and security concerns. Journal of Medical Internet Research, 26, e53008. https://doi.org/10.2196/53008

[3] Russo, D. (2023). Navigating the complexity of generative AI adoption in software engineering. Journal of Systems and Software, 195, 111581. https://doi.org/10.1016/j.jss.2023.111581

[4] Park, J., & Vaswani, A. (2025). Ethical co-design of medical AI: Toward participatory governance. AI & Society, 40(1), 45–60. https://doi.org/10.1007/s00146-025-01642-8

[5] Nguyen, H., Torres, M., & D'Angelo, R. (2025). Cross-sectoral AI governance: Balancing innovation and accountability. Technology in Society, 73, 102274. https://doi.org/10.1016/j.techsoc.2025.102274

Integration of RAG Pipelines in Medical Report Generation

Ivan Lorencin[1], Sandi Baressi Šegota[2], Nikola Anđelić[3], Vedran Mrzljak[4]

1Juraj Dobrila University of Pula, Faculty of Informatics, ilorencin@unipu.hr
2University of Rijeka, Faculty of Engineering, sandi.segota@riteh.uniri.hr
3University of Rijeka, Faculty of Engineering, nikola.andelic@riteh.uniri.hr
4University of Rijeka, Faculty of Engineering, verdant.mrzljak@riteh.uniri.hr

This study explores the effectiveness and practical implications of local Retrieval-Augmented Generation (RAG) systems integrated with various large language models (LLMs) for the automated generation of medical documentation. Comprehensive testing included commercially available and publicly accessible LLMs, such as GPT-4, GPT-4o, LLaMA, DeepSeek, and embedding-based reranking strategies, to thoroughly evaluate their accuracy, consistency, and clinical relevance. Also, commercially available no-code solutions such as NotebookLM are also evaluated. Quantitative analyses employed metrics including the F1 Score, BLEU Score, Recall@K, and Key Information Coverage (KIC), which provided objective measures of precision and completeness (Bernardi & Cimitile, 2024; Weinert & Rauschecker, 2025).

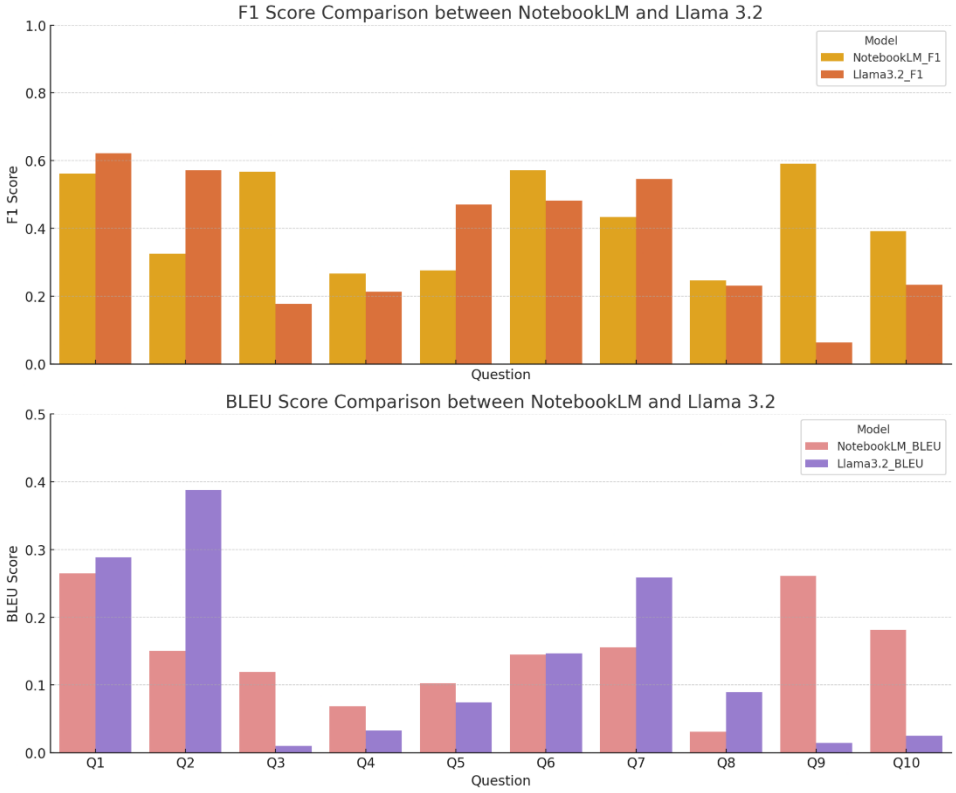


Figure 1 - Comparison of F1 and BLEU scores between NotebookLM and Llama 3.2 across ten medical queries.

Additionally, detailed qualitative assessments through expert reviews validated the clinical applicability of generated documents, emphasizing clarity, diagnostic value, and overall clinical accuracy. The research also examined structured medical data management through advanced automated SQL query generation, revealing significant improvements in precision when embedding-based retrieval methods were combined with reranking techniques (Tanković, Šajina, & Lorencin, 2025). This structured retrieval significantly improved the integration of patient-specific structured data into comprehensive medical documentation. Moreover, practical benefits of locally deployed LLM

solutions were assessed, highlighting their effectiveness in administrative healthcare tasks, secure data management, and regulatory compliance (Lorencin, Tanković, & Etinger, 2025). Local deployment was identified as a key enabler for ensuring data privacy and minimizing risks associated with external cloud-based processing, particularly when handling sensitive patient data. Collectively, the findings from integrating RAG and reranking techniques with structured data retrieval processes demonstrate considerable advancements in reliability, precision, and overall clinical utility of AI-generated medical documentation. When the results are compared, NotebookLM demonstrates superior performance over Llama 3.2 across all evaluated queries, achieving higher F1 scores (e.g., 0.5667 vs. 0.1765 on Q3) and BLEU scores (e.g., 0.2646 vs. 0.0104 on Q1), indicating greater factual accuracy and linguistic quality in the generated medical content, as presented in Figure 1. However, Llama 3.2, as a locally deployable open-source model, offers significant advantages in terms of data privacy and security, making it a compelling choice for privacy-sensitive clinical environments despite its slightly lower performance metrics.

This integrated and locally implemented approach mitigates the limitations of standalone LLM systems by combining retrieval precision with secure, on-premise deployment, thereby enhancing the quality of medical documentation and supporting clinical decision-making while ensuring compliance with stringent data privacy requirements.

Keywords: Retrieval-Augmented Generation, Medical Documentation, Large Language Models, Reranking, Local Deployment, Medical Data Retrieval

References:

[1] Bernardi, M. L., & Cimitile, M. (2024). Report generation from X-ray imaging by retrieval-augmented generation and improved image-text matching. Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN), IEEE, 1–8. https://doi.org/10.1109/IJCNN60899.2024.10650332

[2] Lorencin, I., Tanković, N., & Etinger, D. (2025). Optimizing healthcare efficiency with local large language models. Intelligent Human Systems Integration (IHSI 2025), 160, 576–584. https://doi.org/10.54941/ahfe1005863

[3] Tanković, N., Šajina, R., & Lorencin, I. (2025). Transforming medical data access: The role and challenges of recent language models in SQL query automation. Algorithms, 18(3), 124. https://doi.org/10.3390/a18030124

[4] Weinert, D. A., & Rauschecker, A. M. (2025). Enhancing large language models with retrieval-augmented generation: A radiology-specific approach. Radiology: Artificial Intelligence, 7(3), e240313. https://doi.org/10.1148/ryai.240313

Exploring AI use among elementary school pupils

Davorka Radaković[1], Tanja Savić[2], William Steingartner[3]

[1]University of Novi Sad, Faculty of Sciences, Department Mathematics and Informatics, Novi Sad, Serbia, davorkar@dmi.uns.ac.rs
[2]Elementary School “Nikola Tesla”, Novi Sad, Serbia, tanja.savic@gmail,com
[3]Faculty of Electrical Engineering and Informatics, Technical University of Košice, Košice, Slovakia, william.steingartner@tuke.sk

With the development of digital technologies, the methods of implementing teaching content at all levels of education have changed. Schools are equipped with modern computer equipment, and pupils communicate with teachers and each other online. Also, pupils find solutions to school assignments by searching answers on the Internet.

Artificial Intelligence (AI) applications play a compelling role in everyday activities like adaptive learning, personalized tutoring, and automatic feedback on assignments. In particular, large language models (LLMs) offer new opportunities to introduce semantic foundations ((e.g. variable scope, type correctness, execution models)) when teaching programming—helping pupils understand not just how code works, but why it works, by relating syntax to meaning in a more intuitive and interactive way. Pupils also use AI for helping with homework using step-by-step explanations and resource generation, text summarization, text prediction and autocorrect, language translation, topic research.

AI-powered tools provide: personalised learning experiences, intelligent tutoring, enhancing students experiences, simplify administrative duties for teachers, and enhanced accessibility and inclusivity with assistive technology (Karimi, 2023). The review (Deng 2024) offers important insights for researchers, educators, and policymakers assessing the effectiveness of integrating generative AI into educational practice. In (Lee, 2025) was shown that a scaffolded, knowledge-integration–based Chat Generative Pre-Trained Transformer (ChatGPT) mobile instant messaging system significantly enhances blended learning discussions by leveraging AI’s strengths while addressing its limitations. This has underscored the urgent need to prepare the current generation of young learners with foundational knowledge of AI and to foster informed, responsible attitudes toward its use.

There are many studies (Ilić, 2024) that examine the impact of ChatGPT in higher STEM education, Wang and Fan in (Wang, 2025) indicate that ChatGPT has a large positive impact on improving learning performance. They conducted a meta-analysis of 51 studies to evaluate ChatGPT’s effectiveness in enhancing students’ learning performance, perceptions, and higher-order thinking, of which only one addressed elementary school.

Having all this in mind the aim of this study is to present an analysis of the applications of AI between pupils in the upper grades of elementary school. The research was conducted in several elementary schools in Novi Sad and its surroundings, involving more than 130 pupils. Data was gathered through electronic survey in Google Forms, and quantitative and qualitative analyses were carried out. The following research questions were addressed:

- How did pupils react to AI?
- How did pupils react to AI in education?
- How did AI impact pupils' learning attitudes?
- How did AI affect students' homework performance?
- What are the perceptions of pupils regarding the punishment of academic dishonesty?

Our survey highlights the benefits, challenges, and future prospects of AI in education. In addition, it discusses the perspective of AI to elevate pupils` experiences, and transform teaching and learning methods.

In conclusion, the future impact of AI on higher education in England appears both promising and versatile. Despite the fact that AI has transformed the learning and teaching routines, it also demonstrates various issues that require prompt attention.

Acknowledgements: This work was supported by the project 030TUKE-4/2023 – “Application of new principles in the education of IT specialists in the field of formal languages and compilers”, granted by the Cultural and Education Grant Agency of the Slovak Ministry of Education; ALFA Labs -- trAnsformation pLan For Ai Living Labs, and CZ-1503-06-2526 (Umbrella) -- Development of Computational Thinking.

Keywords: Artificial intelligence, Elementary school, Innovative education, Trustworthy AI. Teaching/learning strategies

References:

[1] Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 105224. https://doi.org/10.1016/j.compedu.2024.105224

[2] Ilić, J., Ivanović,M. & Klašnja-Milićević, A.,(2024) The Impact of ChatGPT on Student Learning Experience in Higher STEM Education: A Systematic Literature Review *2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET)*, Paris, France, pp. 1-9, doi: 10.1109/ITHET61869.2024.10837649.

[3] Karimi, H., & Khawaja, S. (2023). The impact of artificial intelligence on higher education in England. Creative Education, 14(12), 2405–2415. https://doi.org/10.4236/ce.2023.1412154

[4] Lee, H., & Wu, T. (2025). Enhancing Blended Learning Discussions with a Scaffolded Knowledge Integration–Based ChatGPT Mobile Instant Messaging System. Computers & Education, 105375. https://doi.org/10.1016/j.compedu.2025.105375

[5] Wang, J., & Fan, W. (2025). The effect of ChatGPT on students’ learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. Humanities and Social Sciences Communications, 12(1). https://doi.org/10.1057/s41599-025-04787-y

No-Code AI Automation – Data Analysis with Visual Workflow Automation

***Ognjen Ristić*[1]**

[1]Mihajlo Pupin Institute, Belgrade, Serbia, ognjen.m.ristic@pupin.rs

The accelerating adoption of Artificial Intelligence (AI) and automation technologies offers significant potential to improve productivity across engineering disciplines. This paper explores how no-code automation platforms, are used to build sophisticated AI-driven workflows without writing a single line of code. The core mechanism of these platforms will be described, focusing on their visual, modular architecture. The application of no-code AI platforms, like generative design or advanced simulation remains an area with limited documented case studies, their core capabilities present a path for automating crucial data to the engineering workflow. No-code platforms such as Make, Zapier, and Power Automate provide a modular, drag-and-drop environment that lowers the technical barrier to building AI-powered workflows. Their visual interface allows engineers to focus on domain problems rather than programming details.

By connecting and positioning data flow using defined **scenarios** and their constituent **modules**, automation platforms enable significant efficiency gains. The main benefits of these automations are time saving, reduction in manual data handling errors, improvements in data consistency, and increased overall workflow speed. Scenario is a **trigger**, an event that initiates the automation, such as a new result file being uploaded to a designated cloud drive. Following this trigger, a chain of interconnected **modules** performs specific **actions**. Each module is a pre-built connector to an application (e.g., Google Drive, a text parser, OpenAI, Slack). The power of this system lies in the visual **data mapping**, where an engineer can simply click and drag the output data from one module (like an extracted stress value) and place it into an input field of the next module (like a prompt for an AI model).

Practical workflows will be demonstrated in this paper, an automation that triggers on a new file, uses a text-parsing module to extract important information, and then send it to an AI model.

Such example is provided bellow (Fig.1)

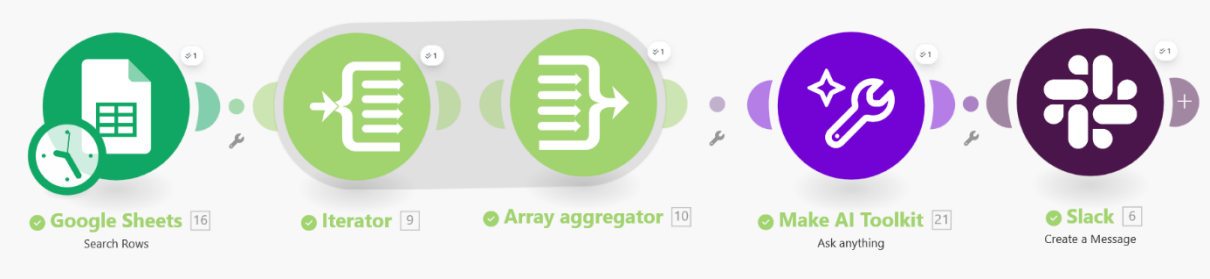


Fig. 1: An example of scenario using make.com

This workflow example is designed to monitor Google Drive for new simulation results, analyze key data using AI, and then report these findings to a Slack channel. The integration leverages Google Sheets, OpenAI's AI models (likely ChatGPT for text analysis), and Slack for communication. This initial step continuously monitors a specified Google Sheet. It specifically looks for rows value. This suggests that certain column likely holds a critical numerical result, such as a stress value, and other column might contain a timestamp or another relevant identifier. The module is configured to extract formatted values and date/time strings. It combines the values from columns and filter rows into an array, which will then be passed to the next module. Next module collects and bundles the data iterated by the Feeder. Data Aggregation aggregates all column values (A-Z) from the filtered Google

Sheet rows, along with metadata like row number, spreadsheet ID, and sheet name. It also includes the combined 'value' from the Feeder module, ensuring that the key filtered data points are carried forward. This is where the artificial intelligence analysis takes place, utilizing a connection to OpenAI (likely a ChatGPT model). The AI receives a detailed prompt instructing it to act as an analyst. It's asked to analyze simulation results, compare them against a defined criterion (specifically referencing the value from certain column of the Google Sheet, which is assumed to be a maximum allowed stress in MPa), and then summarize its findings. The prompt requests the AI to reference the previous result (the maximum stress value).

The AI is prompted to analyze the data, compare it against defined criteria, and generate summary of the results. This summary is then automatically formatted and sent as a real-time notification to the engineering team via a communication platform.

The final step is to forward the analyzed information to a designated Slack channel, then a formatted Slack message is constructed. It includes the file name, value from certain column, the date, other important text and ai reply. The message is sent to a public Slack channel or could alternatively be forwarded with other similar modules.

This workflow creates a robust, automated system for monitoring engineering simulation outputs stored in Google Sheets. It filters for critical results, uses AI to interpret and contextualize these findings against predefined criteria, and then immediately notifies relevant stakeholders via Slack with a concise, AI-generated summary. This significantly streamlines the process of identifying and acting upon important simulation outcomes.

By abstracting the complexity of API calls and scripting, these no-code tools dramatically accelerate the cycle from raw data to the usable results. While the presented workflow demonstrates significant time savings and reduces the likelihood of data handling errors, its reliance on cloud-based services raises considerations of data security and long-term maintainability.

Future developments may include deployment of multimodal AI models capable of analyzing text, numerical, and visual simulation data simultaneously, and ensuring compliance with industry standards.

Keywords: No-Code Platforms, Workflow Automation, Data Processing, Ai automation

References:

[1] Vardanega, T. and Bacchin, F., The Rise of LLM-powered Agents: Reinventing Workflow Automation with Agentic AI., https://platform.aiautomation.cc
[2] Froolik, A.J., 2024. *Effect ai powered email automation: An analysis of email marketing automation* (No. uzsaf_v1). Center for Open Science.
[3] Wynn, M.G. and Skuridin, A., (2025.). AI Assistant for PhD Supervision.
[4] Viswanadhapalli, V., 2025. The Future of Intelligent Automation: How Low-Code/No-Code Platforms are Transforming AI Decisioning. *International Journal Of Engineering And Computer Science*, *14*(1), pp.26803-26825.

A Blockchain-Driven Optimization Mechanism for Federated Learning Data Distribution

Wei Shao[1,2,3], Lianhai Wang[1,2], Qizheng Wang[1,2,3], Jinpeng Wang[1,2], Shujiang Xu[1,2],Shuhui Zhang[1,2]

[1]Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[2]Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, Jinan, China
[3]Key Laboratory of Data and Intelligent System Security, Ministry of Education, Nankai University, Tianjin, 300350, China

Federated learning (FL) enables multiple data holders to collaboratively train a global model without exchanging raw data, thus preserving data privacy. However, in practical deployments, FL systems are often plagued by data heterogeneity—such as imbalanced sample sizes, label distribution skew, and incomplete local datasets—which significantly degrades model convergence, accuracy, and fairness.

To address these challenges, we propose a blockchain-driven hierarchical optimization mechanism for federated learning data distribution that balances efficiency, trust, and privacy. As shown in Fig. 1, our architecture consists of three layers: the Local Client Layer, the Mediator Layer, and the Global Aggregation Layer.

At the base Local Client Layer, we propose a privacy-preserving data sharing mechanism that enables partial, desensitized data or statistical information to be circulated across organizations. Clients compute local data statistics—such as KL divergence in feature and label distributions—and, based on predefined thresholds, are categorized into data-rich, biased, or sparse roles. A skew-aware rebalancing strategy is then triggered to guide the fusion of multi-source data. The use of Blockchain ensures that data provenance and integrity are verifiable without leaking raw content, supporting traceable yet private data admission.

The intermediate Mediator Layer introduces a novel Data Distribution Mediator (DDM) constructed using linear Function Secret Sharing (FSS). Clients encode sensitive data descriptors into function shares and distribute them across non-colluding servers. These mediators cannot access raw values but can collaboratively compute global statistics and make redistribution decisions. To enhance auditability and correctness, we extend FSS to a Verifiable FSS (VFSS) scheme using verifiable MACs. The mediator layer also incorporates smart contracts for bidirectional access control to manage a secure, collaborative data resource pool. Data flow directions are regulated via a dynamic, policy-aware scheduling algorithm embedded within smart contracts.

At the top layer, we design a privacy-preserving model update mechanisms that allow mediators collect masked or encrypted model updates from clients, perform weighted federated averaging under cryptographic protection (e.g., secure aggregation or VFSS-based aggregation), and log all aggregation-related events immutably on the blockchain.

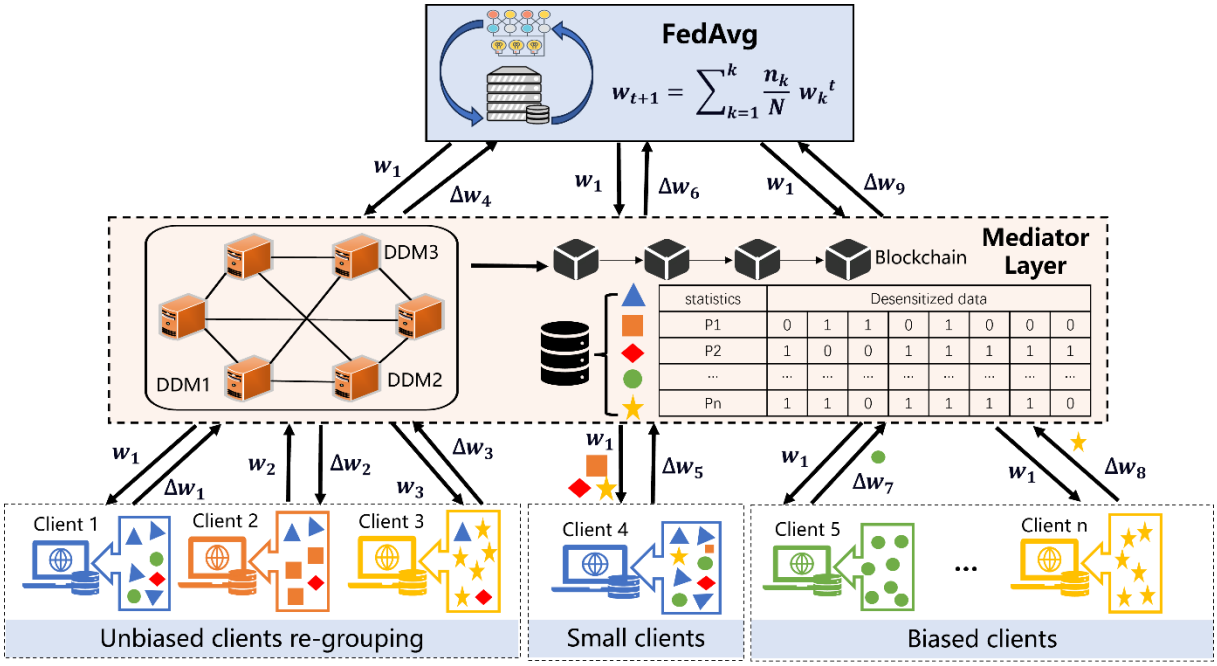


Fig.1 Hierarchical Federated Learning Framework Based on Data Distribution Mediators

Experiments were conducted on benchmark datasets CIFAR-10 under controlled Non-IID configurations. Evaluation metrics included model accuracy, convergence rate, communication overhead, and privacy leakage. Compared with baseline methods such as FedAvg or FedProx, our approach achieved superior accuracy and convergence under heterogeneous data settings. Moreover, privacy leakage was substantially reduced due to cryptographic protection, and communication efficiency was improved via compact representations and decentralized orchestration.

In conclusion, this work presents an integrated solution to the federated learning data distribution challenge, achieving effective collaboration among heterogeneous clients without sacrificing data confidentiality. By leveraging FSS, blockchain, and hierarchical coordination, our method contributes a robust and scalable design for privacy-aware distributed machine learning. Future directions include supporting real-time client updates, adaptive data sampling policies, and applying the system to domains such as healthcare and finance where trust and privacy are paramount.

Keywords: federated learning, blockchain, Non-IID data, secure data sharing

References:

[1] Yu, G., Wang, X., Sun, C., & others. (2025). Ironforge: An open, secure, fair, decentralized federated learning. *IEEE Transactions on Neural Networks and Learning Systems, 36*(1), 354–368.

[2] Boyle, E., Gilboa, N., & Ishai, Y. (2016). Function secret sharing: Improvements and extensions. In T. Holz & S. Savage (Eds.), *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS)* (pp. 1292–1303). ACM Press.

[3] Yang, R., Zhao, T., Yu, F. R., & others. (2024). Blockchain-based federated learning with enhanced privacy and security using homomorphic encryption and reputation. *IEEE Internet of Things Journal, 11*(12), 21674–21688.

[4] Qin, Z., Yan, X., Zhou, M., & others. (2024). BlockDFL: A blockchain-based fully decentralized peer-to-peer federated learning framework. In *Proceedings of the ACM Web Conference 2024* (pp. 2914–2925). ACM.

Game Engine-Based RGB-Thermal Image Simulation: A Platform for Future AI Integration

Bojan Šmidling[1], Miona Tomić[1]

Institute Mihailo Pupin, University of Belgrade, Belgrade, Serbia, bojan.smidling@pupin.rs

We present a software-based thermal image synthesis system that addresses critical limitations in current thermal simulation methodologies. Our approach combines game engine-based 3D thermal modeling with proposed AI enhancement strategies to enable scalable, realistic thermal imagery generation for military training, algorithm development, and search-and-rescue applications.

The implemented simulation system operates within a game engine framework, providing a synthesis of RGB(visible range)–thermal image pairs through a manual assignment of thermal parameters to 3D scene objects, shown in the figure Fig. 1. The system produces thermal signatures by defining emission coefficients, reflectance properties, and temperature distributions across virtual environments. The current implementation focuses on ideal conditions, omitting atmospheric effects such as fog, rain, or humidity. While such conditions allow for clearer validation of simulation principles, future versions should incorporate more complex environmental factors to enhance robustness. The manual approach has proven effective for controlled scenario development and proof-of-concept demonstrations, supporting rapid prototyping of thermal training environments without reliance on thermal imaging hardware.

However, our analysis reveals significant limitations to the manual workflow. Specifying parameters becomes prohibitively labor-intensive when accounting for subtle thermal variations—such as camouflage effects, material-specific heat signatures, dynamic temperature shifts, and environmental influences for each and every object in the virtual environment. The high-dimensional parameter space, particularly under conditions involving occlusion, movement, and atmospheric variability, limits feasibility for large-scale deployment.

To overcome these limitations, we outline the potential of integrating deep learning models capable of autonomously extracting visual and contextual information to assign thermal properties without human input (Hybrid system). Generative architectures such as Conditional GANs (Mirza, 2014), CycleGANs (Zhu, 2017), and multimodal transformers (Lu, 2021) have demonstrated the ability to learn cross-modal mappings and infer thermal behavior patterns from RGB data. Hybrid system's AI model could replicate emission cues, detect equipment heat profiles, and generate object-specific thermal predictions based on learned correlations.

A key obstacle to realizing such AI-based synthesis is the scarcity of large-scale, high-quality RGB–thermal datasets (Shivakumar, 2019). Thermal radiation depends not only on surface appearance but also on intrinsic material and temperature properties, which are not directly observable in visible light. To address this obstacle, we propose leveraging our current simulation system as a synthetic dataset generator. By producing annotated RGB–thermal pairs from controlled scenes, we can create physically grounded data to support future model training—similar to synthetic approaches used in pedestrian re-identification and robotics (Kniaz, 2018).

If described models and datasets are developed, they could enable Hybrid system to synthesize thermal imagery based solely on RGB input. Hybrid system's AI would perform object recognition, segment thermally relevant regions, and synthesize heatmaps informed by both visual and contextual features—without relying on direct thermal measurements.

While we do not currently implement AI model for Hybrid system, our simulation system provides a foundation for future integration. It supports scalable, real-time RGB–thermal synthesis for critical

applications in defense, search and rescue, and autonomous systems. Potential future applications may include drone-based thermal surveillance, intelligent firefighting simulations, and low-cost thermal data augmentation for computer vision research.

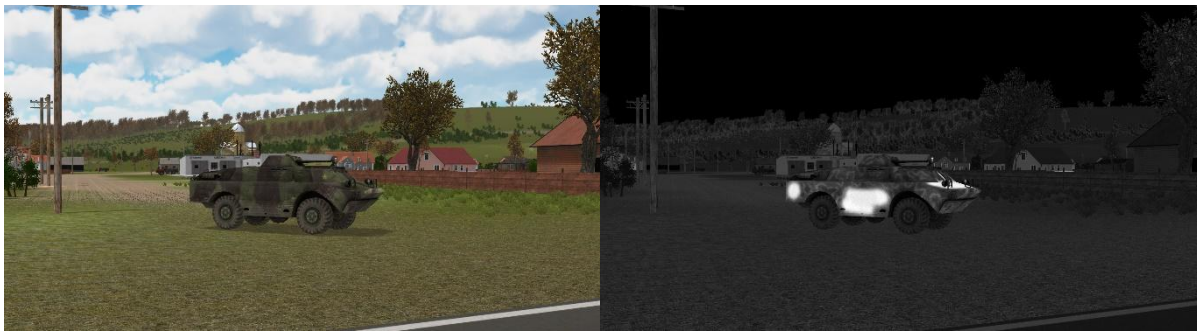


Fig. 1. Screenshot from the simulation software showing a pair of synthetized RGB image (left) and thermal image (right). Thermal image demonstrates visible heat signature originating from vehicle engine and warmed up wheels as a consequence of vehicle movement

Keywords: IR image synthesis, Real time optoelectronics simulation, RGB-thermal imagery, game engine modeling, deep learning integration

References:

[1] Kniaz, V., et al. (2018). ThermalGAN: Multimodal color-to-thermal image translation for person re-identification. *Multispectral Dataset: Munich, Germany, September 8-14, 2018, Proceedings, Part VI*

[2] Lu, K-H., et al. (2021). A Transformer-based Cross-modal Fusion Model with Adversarial Training for VQA Challenge 2021. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.

[3] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

[4] Shivakumar, S. S., et al. (2019). PST900: RGB–thermal calibration, dataset and segmentation network. *2020 IEEE International Conference on Robotics and Automation (ICRA)*

[5] Zhu, J.-Y., et al. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Effects of Investment in the Application of Artificial Intelligence Technologies on the Transformation of the Financial

***Nikola Vidović*[1]**

[1] Faculty of security studies, University of Belgrade,Serbia, vidovicnikola.finance@gmail.com

The rapid advancement of artificial intelligence (AI), particularly generative AI, is reshaping the financial services industry. Driven by the exponential growth of digital data, machine learning, and computational power, financial institutions are increasingly leveraging AI to boost efficiency, foster innovation, and gain competitive advantage.

This study examines the impact of AI investments on the structural transformation of the financial sector, focusing on operational cost reduction, improved decision-making in fraud detection and credit risk assessment, enhanced regulatory compliance, and long-term strategic benefits. Using a mixed-method approach, the research combines quantitative financial performance analysis with qualitative case studies from banking, insurance, auditing, and fintech. Data sources include academic literature, industry reports, and empirical evidence.

The surge in digital payments and online banking has created vast volumes of transactional data, enabling AI to generate insights into consumer behavior and support personalized engagement, often through “next best action” models. A key innovation is the adoption of neuro-genetic algorithms—systems combining neural networks and genetic algorithms—to optimize decision-making, resource allocation, and predictive analytics.

A core argument of the paper is that AI-related return on investment (ROI), especially for generative AI, should not be measured solely by short-term financial gains. Long-term factors—such as talent development, customer and employee experience, sustainability, and responsible AI practices—must be integrated into evaluation frameworks. Evidence suggests that systematic AI investment reduces costs and enhances resilience, adaptability, and innovation.

In today’s dynamic financial landscape, ongoing innovation is essential for institutional survival. Generative AI is projected to increase labor productivity by 0.1% to 0.6% annually through 2040, depending on adoption rates and workforce reallocation. Combined with automation and other technologies, productivity growth could rise by 0.5 to 3.4 percentage points annually.

Realizing these benefits requires workforce upskilling and agile strategies to manage transition risks. With effective governance, generative AI could contribute to inclusive economic growth and support a sustainable, human-centric financial system.

This paper offers a comprehensive, evidence-based analysis of AI’s transformative potential in finance and presents strategic priorities for scaling its benefits, providing guidance for industry leaders and policymakers focused on long-term value creation.

Keywords: Financial sector, artificial intelligence, transformation expenses, return on investment, digital economy.

References:

[1] Azagury, J., Ashraf, M., Wright, O., Fang Grant, K., Moore, M. (2024). Reinvetion in the age of generative AI, Research Report, pp. 1-53, Accenture. Available at: www.accenture.com/research.

[2] Becerra-Vicario, R., Salas-Compás, B., Valcarce-Ruiz, L., & Sánchez-Serrano, J. R. (2024) THE IMPACT OF ARTIFICIAL INTELLIGENCE IN THE FINANCIAL SECTOR: OPPORTUNITIES

AND CHALLENGES. International Journal of Business & Management Studies, Vol. 05, Issue No 10.
[3] Vučenović, S., Vukmirović, G., Marić, R., Macura, N. (2024). The potential application of artificial intelligence in retail business in the shadow of the controversy around technology. Economic Ideas and Practice 55, pp. 1-12. Available at: https://doi.org/10.54318/eip.2024.sv.382.
[4] World Economic Forum (2025). Artificial Intelligence in Financial Sevices. Available at: https://reports.weforum.org/docs/WEF_Artificial_Intelligence_in_Financial_Services_2025.pdf
[5] Webb, A. (2019). The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity, pp. 1-320, Public Affairs, New York.

Efficient Feature Disentanglement via Orthogonal Decomposition for Model Fine-tuning

Jiahui Yang [1], Jianxun Cui [1]

[1]School of Transportation Science and Engineering, Harbin Institute of Technology, 23S136655@stu.hit.edu.cn

Effective feature disentanglement and control in large-scale generative model fine-tuning remains a significant challenge. This paper presents a novel low-rank fine-tuning method based on QR decomposition [1], achieving efficient separation and control of content and style features.

Our method is built upon a key insight: decomposing the traditional LoRA's [2] low-rank matrix into an orthogonal matrix Q and an upper triangular matrix R, where Q serves as a fixed orthogonal basis set while R is adjusted through a learnable increment ΔR. This design enables features from different tasks (such as content and style) to be projected onto mutually orthogonal subspaces, thereby achieving better feature disentanglement. Specifically, for each weight matrix W that needs fine-tuning, we first decompose it into $W = QR$, then introduce a learnable increment matrix ΔR, making the fine-tuned weight $W + \Delta W = Q(R + \Delta R)$. This design not only reduces trainable parameters by half but also provides better feature separation capability through orthogonal basis constraints.

In our experiments, we validate the effectiveness of this method across multiple mainstream text-to-image models [3,4,5]. Compared to traditional LoRA approaches, our method achieves faster convergence and better feature disentanglement with fewer parameters. Quantitative analysis shows that the cosine similarity between feature vectors of different tasks significantly decreases under orthogonal constraints, confirming improved feature separation. In qualitative analysis, we demonstrate that our method better preserves target content while achieving precise control over style features.

Furthermore, our method offers several advantages:

1) accelerated model convergence through good initialization provided by QR decomposition;

2) reduced interference between features due to the orthogonal basis set, making the combination of multiple LoRAs more stable;

3) model-agnostic design that can be widely applied to different model architectures.

In summary, our proposed method introduces orthogonal decomposition to achieve more efficient feature disentanglement and control while maintaining model performance. This approach provides a new perspective for customizing large-scale generative models and demonstrates significant practical value.

Keywords: LoRA, Efficient Feature Disentanglement, Image Customized Generation

References:

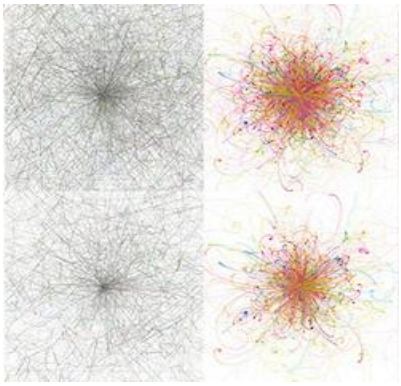
[1] Gene H Golub and Charles F Van Loan. Matrix computations johns hopkins university press. Baltimore and London, 41: 62, 1996. 2

[2] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685, 2021. 2, 4

[3] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Mu¨ller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion mod- els for high-resolution image synthesis. arXiv preprint arXiv:2307.01952, 2023. 2, 3, 6, 7

[4] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Mu¨ller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling recti- fied flow transformers for high-resolution image synthesis. In Forty-first International Conference on Machine Learning, 2024. 2, 3, 6, 7
[5] Black Forest Labs. Flux. https : / / github . com / 571 black-forest-labs/flux, 2023. 2, 3, 6, 7

KNOWLEDGE REPRESENTATION, UNCERTAIN KNOWLEDGE AND REASONING



Hybrid Knowledge Graph and Machine Learning Framework for Code-mixed Anti-Religious Content Detection

Mahima Choudhary[1a], Basant Agarwal[2a]

[a] Department of Computer Science and Engineering, Central University of Rajasthan, Kishangarh, Ajmer, India, [1]mahimacs23@gmail.com, [2]basant@curaj.ac.in

The proliferation of social media platforms has created significant challenges for detecting anti-religious hate speech, particularly in code-mixed hindi-english content. This research introduces Code-Mixed Knowledge Graph, a novel knowledge graph architecture specifically designed for multilingual religious hate speech detection.

Previous research in hate speech detection has evolved from lexicon-based approaches to transformer-based models and large language models (LLMs) (Davidson et. al. 2017). However, these approaches often fail to capture complex semantic relationships between religious entities, cultural practices, and hate speech terminology in code-mixed environments where speakers seamlessly switch between languages, combining English with local languages such as Hindi (Founta et. al. 2018).

This study presents a comprehensive methodology for constructing and utilizing a code-mixed knowledge graph for hate speech detection, with particular emphasis on comparing two fundamental embedding aggregation strategies: averaging entity embeddings versus concatenating them. The research addresses three critical questions: (1) the feasibility of constructing knowledge graphs for code-mixed structured data, (2) the comparative impact of different embedding aggregation techniques on hate speech detection performance, and (3) the identification of machine learning classifiers that benefit most from concatenated entity embeddings.

The Code-Mixed Knowledge Graph construction framework integrates entity extraction, relationship mapping, and embedding generation to create robust representations of religious hate speech patterns. The knowledge graph encompasses eight distinct entity categories with comprehensive multilingual support: Religion, Religious_Text, Religious_Figures, Religious_Practices, Religious_Concepts, Religious_Places, Religious_Groups, and Hate_Speech terms. Each category incorporates religious terms in multiple languages, including English and Hindi, with Romanized and Devanagari scripts.

Entity embeddings are derived from MuRIL (Multilingual Representations for Indian Languages) transformer models (Chen et. al. 2024). The knowledge graph models two primary relationship types: RELATED_TO (direct semantic relationships between entities and target religions) and CO_OCCURS_WITH (co-occurrence relationships between entities within textual content). Each relationship is annotated with sentiment polarity categorized as negative, positive, mixed, or neutral contexts.

Experimental evaluation was conducted using the THAR (Targeted Hate Against Religion) dataset, comprising 11,549 sentences divided into AntiReligion and Non-AntiReligion classes. The dataset was partitioned using an 80:10:10 ratio for training, validation, and testing. Performance evaluation employed standard metrics across three machine learning classifiers: Naive Bayes, Logistic Regression, and Support Vector Machine.

The constructed Code-Mixed Knowledge Graph incorporates more than 60% of entities being non-English terms, authenticating its multilingual coverage. Experimental results demonstrate that concatenated entity embeddings consistently outperform averaged embeddings across all classifiers.

The superior performance of concatenated embeddings can be attributed to their ability to preserve entity-specific semantic information. This finding suggests that preserving granular entity information is crucial for effective semantic understanding in complex domains such as hate speech detection.

Future research directions include exploring sophisticated aggregation techniques beyond concatenation and averaging, implementing automated systems for detecting emerging hate speech terminology, and extending the methodology to create universal knowledge graphs covering multiple languages and cultural contexts globally.

Keywords: Hate Speech Detection, Knowledge Graph, Code-mixed, Embeddings, Anti-Religious Content

References:

[1] Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. ACM Computing Surveys, 51(4), 1-30.
[2] Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Hate speech detection with a convolution neural network. In Proceedings of the 11th International AAAI Conference on Web and Social Media (pp. 512-515).
[3] Founta, A., Djouvas, C., Chatzakou, D., Leontiadis, I., Blackburn, J., Stringhini, G., ... & Kourtellis, N. (2018). Large scale crowdsourcing and characterization of twitter abusive behavior. In Proceedings of the International AAAI Conference on Web and Social Media (pp. 491-500).
[4] Chen, M., et al. (2024). A knowledge graph embedding model based attention mechanism for enhanced node information integration. PeerJ Computer Science, 10, e1910.
[5] Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. IEEE Transactions on Knowledge and Data Engineering, 29(12), 2724-2743.

Complete meet-continuous codomain lattice and extremal solutions of L-valued fuzzy relation equations

Vanja Stepanović[1], Andreja Tepavčević[2,3]

[1]Faculty of Agriculture, University of Belgrade, Serbia, vanja@agrif.bg.ac.rs
[2]Mathematical Institute, Serbian Academy of Sciences and Arts, Belgrade, Serbia, andreja@dmi.uns.ac.rs
[3]Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Serbia, andreja@dmi.uns.ac.rs

Fuzzy relational equations are essential components of fuzzy control systems, enabling the implementation of rule-based logic in uncertain environments. In a typical Mamdani-style controller, each fuzzy IF–THEN rule is represented as a fuzzy relation, and the overall rule base is fused by composing these relations with the current input membership grades. Solving the resulting fuzzy relational equations yields an aggregated output fuzzy set that can then be defuzzified into an exact control action. Because they naturally handle uncertainty and partial truth, fuzzy relational equations allow controllers to operate robustly in environments where classical, precise models are impractical or unavailable.

In our work, we use a more general notion of fuzzy relation, so-called lattice-valued fuzzy relation, which maps pairs of elements of the domain set into a lattice (instead of [0,1] interval). We use sup-inf fuzzy relational composition, which is well-defined in the case of a complete codomain lattice, and also order-preserving relative to the component-wise ordering. Relative to that ordering, the set of fuzzy relations is a lattice. Relative to the composition, the set of fuzzy relations is a groupoid, another algebraic structure that can be studied by itself.

When speaking about extremal solutions to fuzzy relational equations, the case of a complete codomain lattice is often too general. For example, when a typical equation is concerned, i.e., an equation of the type $P \circ Q = R$, or $Q \circ P = R$ where P and R are known fuzzy relations, Q is an unknown fuzzy relation; there might be neither a maximal nor a minimal solution, which makes an optimization task undefinable.

That's why many investigators tend to add some additional conditions in order to get some more regularity in the set of solutions (Diaz-Moreno et al., 2017; di Nola & Lettieri, 1989). We add the property of meet-continuity of the codomain lattice, which means that the infimum commutes with the supremum of any chain. This is a natural condition, fulfilled by many lattices, among which are all algebraic lattices, i.e., lattices derived from an algebra, such as the lattice of all congruences, or the lattice of all weak congruences of an algebra. Such a condition is weaker than infinite distributivity of infimum relative to supremum, fulfilled in some more special lattices, such as Browerian lattice.

Under the condition of meet-continuity, the existence of a maximal solution was proved for some fuzzy relational inequations and equations already in 2018 (Stepanović, 2018). To prove the existence of maximal solutions to some other fuzzy relational equations and inequations, such as $Q \circ P = Q$, or $P \circ Q = Q$, a property of "composition-continuity" in the set of relations was proved later on. This means that the composition, both from the left and from the right-hand side, commutes with the supremum of chains. A more general property is needed and proved in order to prove that, e.g., equations like $Q \circ Q = P$ or $Q \circ Q = Q$ have a maximal solution. We prove that, given a chain of pairs of fuzzy relations (P_i, Q_i) (relative to the componentwise ordering) satisfy that

$$\bigvee_{i\in I} P_i \circ \bigvee_{i\in I} Q_i = \bigvee_{i\in I} P_i \circ Q_i \qquad (1)$$

In order to get more general results, we form terms using variables, constant fuzzy relations and, as operations, fuzzy relational composition, supremum, and infimum. Now, the property (1) implies, by induction on the complexity of terms, that for a given term $T(X_1, \dots, X_n)$, and a set of chains of n-tuples $(Y_1^i, \dots, Y_n^i)$, the function of n variables defined by the term commutes with the supremum of the chain, i.e., we have that

$$T\left(\bigvee_{i\in I} Y_1^i, \dots, \bigvee_{i\in I} Y_n^i\right) = \bigvee_{i\in I} T(Y_1^i, \dots, Y_n^i) \qquad (2)$$

Using (2), we prove that every equation of the type $T_1(X_1,,,,,X_n) = T_2(X_1,,,,,X_n)$ has a set of solutions that is closed under the supremum of chains. This further implies that every non-empty set of solutions has an upper bound, i.e., its supremum. Applying Zorn's lemma, we conclude that there exists a maximal solution.

Actually, we have proved that, basically, any fuzzy relational equation in the case of a meet-continuous codomain lattice has a maximal solution. A question naturally arising is whether there exists a greatest solution as well. This needn't be, as was shown by a counterexample (Stepanović, 2018). However, certain types of equations do have the greatest solution, even without the condition of meet-continuity. (Stepanović & Tepavcević, 2022). There also need not exist a minimal solution. We provide an example proving this.

Acknowledgements: This research was supported by the Science Fund of the Republic of Serbia, Grant no 6565, Advanced Techniques of Mathematical Aggregation and Approximative Equations Solving in Digital Operational Research- AT-MATADOR. The authors also gratefully acknowledge the financial support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Grants No. 451-03-137/2025-03/200116, 451-03-137/2025-03/200125 & 451-03-136/2025-03/200125 & 451-03-136/2025-03/200029)

Keywords: meet-continuous lattice, general equation, maximal solution, greatest solution, minimal solution.

Refernces:

[1] Díaz-Moreno, J.C., Medina, J., Turunen, E. (2017). *Minimal solutions of general fuzzy relation equations on linear carriers. An algebraic characterization*, Fuzzy Sets and Systems, vol. 311, pp. 112-123.

[2] di Nola, A., Lettieri, A. (1989). *Relation equations in residuated lattices*. Rend. Circ. Mat. Palermo, vol. 38, pp. 246–256.

[3] Stepanović, V. (2018). *Fuzzy set inequations and equations with a meet-continuous codomain lattice*, Journal of Intelligent & Fuzzy Systems, vol. 34/6, pp. 4009-4021.

[4] Stepanović, V., Tepavčević, A. (2022). *Fuzzy Sets (In)Equations with a Complete Codomain Lattice,* Kibernetika, vol. 58/2, pp. 145-162.

Hesitant fuzzy sets and hesitant fuzzy relations over the Gödel power set quantale

Marija Tasić[1], Miroslav Ćirić[2], Jelena Ignjatović[2], Marko Stanković[1], Stefan Stanimirović[2]

[1]Pedagogical Faculty in Vranje, University of Niš, Serbia, marijat@pfvr.ni.ac.rs, markos@pfvr.ni.ac.rs
[2]Faculty of Sciences and Mathematics, University of Niš, Serbia, mciric@pmf.ni.ac.rs, jelena.ignjatovic@pmf.ni.ac.rs, stefan.stanimirovic@pmf.ni.ac.rs

Traditional quantitative methods of systems analysis are inherently inadequate for dealing with systems that are based on human reasoning or behavior. Given that much of human reasoning is grounded in vague truth values, approximate logical operations, and imprecise inference, Lotfi A. Zadeh proposed an alternative approach based on the notion that core components of human cognition are fuzzy categories of objects, characterized by a smooth transition between membership and non-membership (Zadeh, 1965). Fuzzy set theory, since its origin, has played a significant role in formalizing aspects of human reasoning and knowledge representation.

As research has progressed, it has become evident that certain decision-making systems require more diverse and flexible frameworks to operate effectively, leading to the development of various extensions of classical fuzzy set theory. Among these, hesitant fuzzy set theory has been widely studied as an approach that offers effective tools for modeling problems that involve hesitation, uncertainty, and vague or imprecise information.

Hesitant fuzzy sets, introduced by Torra and Narukawa (Torra & Narukawa, 2009) in 2009, extend fuzzy sets by allowing membership degrees to be represented as subsets of the real unit interval. This framework captures uncertainty in decision-making by representing the membership of an element as a set of possible values, rather than a single degree, thereby accounting for multiple assessments from different decision-makers. This aspect makes hesitant fuzzy sets particularly suitable for problems where decision-makers are unsure or express multiple preferences. Therefore, hesitant fuzzy sets have a significant role in artificial intelligence and machine learning, especially in areas such as knowledge representation, natural language processing, and decision-making systems that deal with uncertainty and vagueness.

However, it is well-known that Torra's operations do not produce a lattice and one of the important open problems in this area is to construct a lattice structure on hesitant fuzzy sets that preserves Zadeh's meet and join operations when restricted to classical fuzzy sets. Jara et al. addressed this problem and, in their work (Jara, Merino, Navarro, & Santos, 2023), proposed a definition of a symmetric order on hesitant fuzzy sets, which resulted in a lattice structure. Nevertheless, the corresponding symmetric meet and join operations do not coincide with those defined by Torra.

In this work, we propose a novel approach to hesitant fuzzy sets by defining them over the power set quantale $\mathcal{P}(S)$ of an arbitrary semigroup S. We have introduced specific hesitant fuzzy sets that form a bounded lattice under set inclusion. Additionally, a new operation based on the Gödel's conjunction in fuzzy logic has been defined here. This approach led us to the construction of a power set quantale over hesitant fuzzy sets, termed the Gödel power set quantale. By doing so, we have introduced the residuals for hesitant fuzzy sets.

As fuzzy relations provide a fundamental role in handling uncertain or imprecise data, we extended our framework to hesitant fuzzy relations. Following the approach of Ćirić et al. (Ćirić, Ignjatović, & Bogdanović, 2007) and Ignjatović et al. (Ignjatović, Ćirić, Šešelja, & Tepavčević, 2015), we have provided the characterization of hesitant fuzzy relations, with particular focus on hesitant fuzzy equivalences, classes, and partitions. Our results ensure new insights into the algebraic and relational

aspects of hesitant fuzzy sets, expanding their applicability in uncertain reasoning and decision analysis.

Keywords: Hesitant fuzzy sets, Power set quantale, Hesitant fuzzy relations, Hesitant fuzzy equivalences

References:

[1] Ćirić, M., Ignjatović, J., & Bogdanović, S. (2007). Fuzzy equivalence relations and their equivalence classes. *Fuzzy Sets and Systems*, *158*(12), 1295–1313. https://doi.org/10.1016/j.fss.2007.01.010

[2] Ignjatović, J., Ćirić, M., Šešelja, B., & Tepavčević, A. (2015). Fuzzy relational inequalities and equations, fuzzy quasi-orders, closures and openings of fuzzy sets. *Fuzzy Sets and Systems*, *260*, 1–24. https://doi.org/10.1016/j.fss.2014.05.006

[3] Jara, P., Merino, L., Navarro, G., & Santos, E. (2023). A lattice structure on hesitant fuzzy sets. *IEEE Transactions on Fuzzy Systems: A Publication of the IEEE Neural Networks Council*, *31*(6), 2018–2028. https://doi.org/10.1109/tfuzz.2022.3217400

[4] Torra, V., & Narukawa, Y. (2009). On hesitant fuzzy sets and decision. *2009 IEEE International Conference on Fuzzy Systems*. https://doi.org/10.1109/FUZZY.2009.5276884

[5] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/s0019-9958(65)90241-x

Intelligent data – the platform driven approach

Milorad Tosic[1], Nenad Petrovic[2]

[1]Faculty of Electronic Engineering, University of Nis, Serbia, milorad.tosic@elfak.ni.ac.rs
[2]Faculty of Electronic Engineering, University of Nis, Serbia, nenad.petrovic@elfak.ni.ac.rs

The question “How intelligent is an intelligent system?” (Devedžić, 2020) has been guiding research in the field of Artificial Intelligence (AI) for yeas so far. Now, let us make the next step and ask “Where the intelligence is coming from?” Intelligent system is a system consisting of interdependent collaborating parts that together constitute a new, more complex and more valuable, entity. Based on literature (Sternberg, 1987), intelligence has at least following subthemes: knowledge, learning, decision-making, actions/communication. Intelligence is a process, not a set characteristic. The main proposition of this paper is that *intelligence cannot exist without communication between parts of the intelligent system*. In the system without communication the question “How intelligent is Schrödinger's cat?” is senseless (Figure 2a) and consequently any form of intelligence is impossible. On the next level, we have Searle’s Chinese Room argument (Figure 2b), where communication between actors

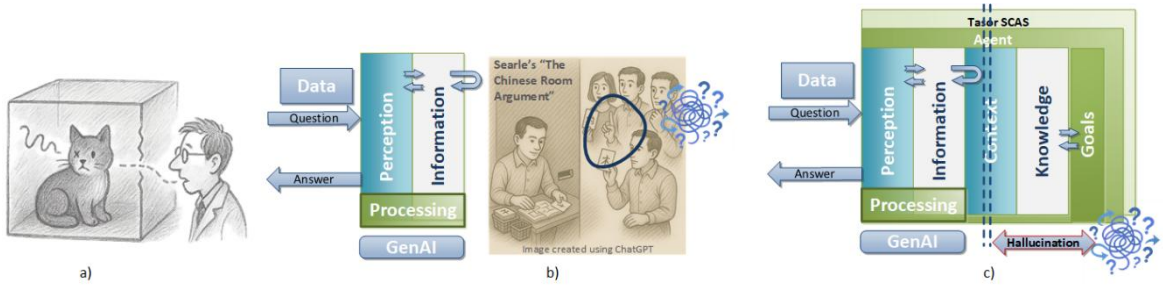


Figure 2. Simulated intelligence. a) No communication - no intelligence. b) Intelligence at processing level - Chinese room argument. c) Explanation of LLM's "hallucinations"

exists, as translated output is generated based on input notes. However, the overall interaction lacks context and understanding, despite that the room leaves impression of intelligent behavior based on generated outputs by applying simple mechanics against the input data. It indicates that the intelligence of LLMs should be located at this level. Note the “hallucinated” hand on the illustration image that is generated using ChatGPT. At the information level, the hand on the image is perfectly fine even though it does not make sense because it is not attached to any of the present persons. Figure 2c explains that the LLMs’ “hallucinations” are direct consequence of missing context and further lack of shared knowledge that results in the inability to understand goals. Figure 3 introduces the concept of Intelligent Data (InD) where plain data is integrated with semantics. The InD based communication facilitates contextualized communication of data based on shared knowledge. In this way, system is capable to reason and thanks to that to be aware and understand goals (and eliminate hallucinations). The next development stage in intelligent systems, often called Super Intelligence (SuI), is illustrated in Figure 3b. In addition to InD, SuI would include intelligent actions powered by some kind of meta-reasoning.

In this paper, we propose the following extension of traditional semantics-driven systems paradigm with the goal to implement InD. Plain Content (PC) is traditional data associated with URL that is interpreted as URI in the semantic environment. Reference to the URI of the PC retrieves the content. Semantic resource that represents meaning of the PC is connected to the URL of the PC by property td:hasPlainContent (the property is defined in the TasorData ontology defining all aspects of InD). This way, we introduce the extension of standard RDF language, providing the support for URL as object within semantic triplets, apart from URI, String and BlankNode. In this context, this extension represents URL of file which contains plain data. td:hasPlainContent property gives us the ability to link plain data

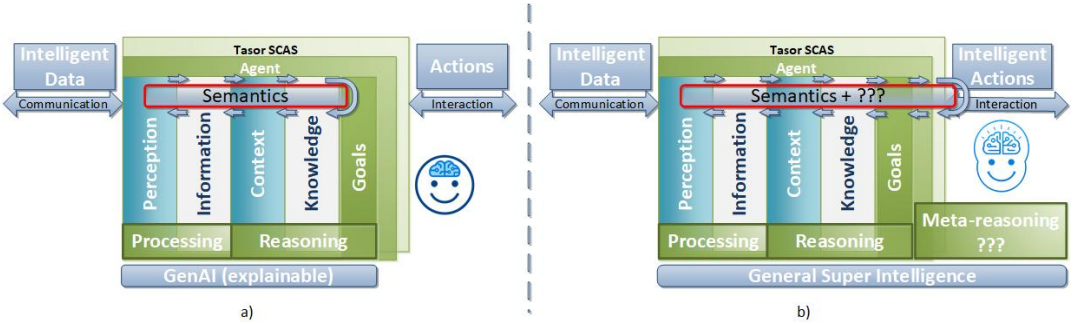


Figure 3. Intelligent Data - from Deep (a) to Super Intelligence (b)

with their semantic description as a whole. TasorSCAS semantic platform[3] is developed that implements the extension. Note that implementation of the concept is possible using the platform approach only. To the best of our knowledge, there is no such existing approach that bridges the gap of connecting plain content with its semantics. This way, we introduce additional value to plain data, going step towards InD, opening new horizon of applications, such as enabling reasoning capabilities within traditional data management platforms. Adoption of InD within organizations has many advantages, such as shared understanding across various business functions and domains, making new use cases possible, as well as integration and interoperability more convenient (Figure 4). It is promising for software intensive companies where reducing cognitive load of administrators and giving the ability of managing hundreds of microservices thanks to InD adoption are of crucial importance (Tosic, 2024). The implementation platform has been successfully adopted for low-code development of industry-oriented applications relying on microservices, as well as fatigue monitoring and workforce planning (Tosic, 2024).

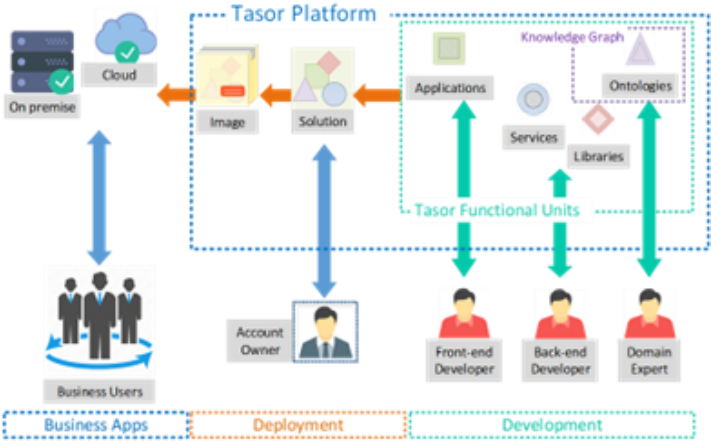


Figure 4. Integrated full-life cycle of intelligent data applications

Keywords: Ontology, meaning, data management, LLM

References:

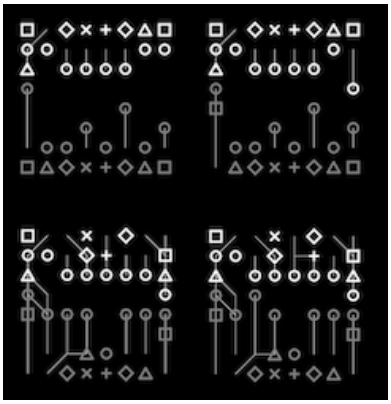
[1] Devedžić, V. (2020). Is this artificial intelligence?. *Facta Universitatis, Series: Electronics and* Energetics, 33(4), 499-529.

[2] Sternberg, R. J. (1987). Intelligence. In R. L. Gregory & O. L. Zangwill (Eds.), (2nd ed., pp. 375–379). Oxford University Press.

[3] Tosic, M., Petrovic, N., & Tosic, O. (2024). Workforce wellbeing management leveraging semantic knowledge graph. In Proceedings of the XX International May Conference on Strategic Management – IMCSM24 (Vol. 20, Iss. 1, pp. 418–427). Technical Faculty in Bor.

[3] www.virtuonasoft.com/scas

MACHINE LEARNING



Swarm Intelligence Methods in Feature Selection for Biomedical Data Classification

Irfan Fetahović[1], Aldina Avdić[1], Ulfeta Marovac[1]

[1]Department of Technical-Technological Sciences, State University of Novi Pazar, Vuka Karadžića bb, Serbia, [ifetahovic, apljaskovic, umarovac]@np.ac.rs

Applying machine learning methods to large datasets with numerous features presents challenges related to training time, model complexity, and interpretability. Feature selection addresses these issues by reducing data dimensionality, improving classification accuracy, and enhancing interpretability [Al-Rajab et al., 2021]. This study aims to improve the classification of integrated biomedical data for thrombophilia diagnosis.

The dataset comprises 71 features collected from 35 women, including 22 healthy individuals and 13 diagnosed with thrombophilia. To evaluate model performance, three classification algorithms were applied: K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM). Subjects were recruited from the Gynecology-Obstetrics Clinic of the University Clinical Center Kragujevac, encompassing both healthy pregnant women (control group) and patients with thrombophilia confirmed by molecular diagnostics (FV Leiden, FII, MTHFR, and PAI-1). The dataset includes 37 biomedical attributes obtained from anamnesis, laboratory tests, and ultrasound examinations, along with 34 demographic and lifestyle attributes collected through structured questionnaires.

Key features related to thrombophilia diagnosis were identified using wrapper methods based on swarm intelligence (SI) algorithms. Three algorithms were examined: Ant Lion Optimizer (ALO), Bat Algorithm (BA), and Gray Wolf Optimizer (GWO) [Hambali et al., 2022; Meenachi & Ramakrishnan, 2021; Niu et al., 2021]. Each was combined with two classifiers commonly used in biomedical classification: SVM and KNN. The F1 score was used as the evaluation metric, and each algorithm was run 30 times.

For biomedical data, the Bat Algorithm (BA) achieved the best performance, reaching 0.97 accuracy with only four selected features. In contrast, ALO required more than 30 features to obtain lower results (≤0.86). For demographic data, BA again outperformed other methods, achieving 0.83 accuracy with just two features, compared to weaker results obtained by ALO (≈0.74 with over 50 features) and GWO (≤0.75 with 1–2 features). A baseline scenario with all features included was also evaluated, yielding average accuracies of 0.71 for demographic data (34 features) and 0.68 for biomedical data (37 features). These values were consistently surpassed by swarm intelligence–based feature selection, with the BA–SVM combination reaching 0.97 accuracy using only four features. This confirms that dimensionality reduction not only simplifies models but also enhances predictive performance.

Overall, the results indicate that swarm intelligence-based wrapper methods substantially improve model performance and contribute to more efficient data management, which has important practical implications for thrombophilia diagnostics. In particular, the Bat Algorithm proved highly effective in selecting relevant features from both demographic and biomedical data of pregnant women. A key innovation of this study is the integration of clinical and demographic data, which improved predictive accuracy despite the limited sample size [Marovac et al., 2023]. While the small dataset constrains generalizability, the findings provide a strong foundation for future research. Further work will involve training models on larger datasets and exploring advanced metaheuristic algorithms to achieve more robust performance. The proposed method shows promise for real-world application in thrombophilia detection, reducing the need for extensive data collection while maintaining high predictive accuracy.

Keywords: feature selection, machine learning, biomedical data classification, swarm intelligence

References:

[1] Al-Rajab, M., Lu, J., & Xu, Q. (2021). *A framework model using multifilter feature selection to enhance colon cancer classification*. PLOS ONE, 16(4), e0249094. https://doi.org/10.1371/journal.pone.0249094

[2] Hambali, M. A., Oladele, T. O., Adewole, K. S., Sangaiah, A. K., & Gao, W. (2022). *Feature selection and computational optimization in high-dimensional microarray cancer datasets via InfoGain-modified bat algorithm.* Multimedia Tools and Applications, 81(25), 36505–36549. https://doi.org/10.1007/s11042-022-12458-6

[3] Marovac, U., Memić, L., Avdić, A., Djordjević, N., Dolićanin, Z., & Babić, G. (2023). *Selecting critical features for biomedical data classification.* In 2nd International Conference on Chemoinformatics and Bioinformatics (pp. 31–37). University of Kragujevac, Institute for Information Technologies.

[4] Meenachi, L., & Ramakrishnan, S. (2021). *Metaheuristic search-based feature selection methods for classification of cancer*. Pattern Recognition, 119, 108079. https://doi.org/10.1016/j.patcog.2021.108079

[5] Niu, B., Yi, W., Tan, L., Geng, S., & Wang, H. (2021). *A multi-objective feature selection method based on bacterial foraging optimization*. Natural Computing, 20, 63–76. https://doi.org/10.1007/s11047-020-09796-y

CNN-based recognition and correction of axes misalignments and tilts in the wireless communications system utilizing OAM modes for data transmission

Andjelija Ž. Ilić[1], Jelena Z. Trajković[1,2], Slobodan V. Savić[3], Darko M. Ninković[3], Milan M. Ilić[3]

[1]Institute of Physics Belgrade, University of Belgrade, Serbia, andjelijailic@ieee.org
[2]Faculty of Physics, University of Belgrade, Serbia, jelena.trajkovic.ff@gmail.com
[3]School of Electrical Engineering, University of Belgrade, Serbia, ssavic@etf.bg.ac.rs, darko@etf.bg.ac.rs, milanilic@etf.bg.ac.rs

The plans for the beyond-5G / 6G wireless communications systems envision the ubiquitous connectivity of users and devices, with a significant portion of overall data transfer occurring within the machine-to-person (M2P) and machine-to-machine (M2M) communications (Chen *et al*., 2020; Tariq *et al*., 2020). Many of those application scenarios require high-data-rate short-range line-of-sight (LoS) data transmission (immersive extended reality (IXR), industry 4.0, wearable antennas within the body-area networks). In achieving the multigigabit-per-second data rates at a reasonable cost, the OAM wave-based mode-division multiplexing (OAM-MDM) is seen as a low-complexity but high-spectral-efficiency physical layer solution for increasing the data rates. It exploits the orthogonality of OAM modes to transfer multiple data streams over the same channel, thereby contributing to the high communication capacity. It can also be combined with other multiplexing techniques.

Amongst the problems in the practical realization of OAM-MDM, errors due to the misalignments (axes lateral offsets), and tilts of the transmitter and receiver, represent hindering factors in the deployment of OAM wave-based systems, decreasing the quality of the OAM EM field representation in the points of interest and consequently, affecting the system capacity and other figures of merit (Yin *et al*., 2017). Reducing the sensitivity of OAM technology to such errors increases the system robustness and allows maximization of the data rates and system capacity. Since the system adjustments should preferably be performed during the system operation, without incurring significant latency, corrections based on machine learning could provide an efficient, yet sufficiently accurate, solution to the problem.

The OAM EM modes are characterized by the distinct magnitude and phase near-field patterns in the planes perpendicular to the axis of propagation, as illustrated by the plots of instantaneous electric fields of the first four OAM modes in Figure 1. The results shown correspond to the 60 GHz frequency, i.e. 5 mm wavelength in the free space, with the OAM source radii adjusted so as to obtain the maximal effective value of EM field at the radius of 20 mm, viewed at a distance of 120 mm from the source. Electric field points in-phase and out-of-phase with each other are seen in the plots, corresponding to a phase change from zero to $2\pi\ell$ in an observation plane, for the ℓ-th OAM mode. Points with a maximal effective value of EM field correspond to the circular ring patterns in observation planes perpendicular to the propagation axis. In the millimeter-wave frequency region, one of the most convenient and cost-effective ways of generating the OAM EM modes is by the utilization of the uniform circular antenna arrays (UCA), or concentric arrangements of UCA, as explained in (Ilić, Vojnović, *et al*., 2023). The UCA offers relatively easy signal modulation and mode reconfiguration, as well as the straightforward integration with the existing broadband wireless infrastructure. The receiving antenna array is most often also a UCA, picking up the EM fields at discrete locations in the observation plane, and the receiver uses such individual antenna waveforms and performs the field convolutions with the OAM patterns, which in this case are the stream separation factors for the receiver (Ilić, Vojnović, *et al*., 2023). The described procedure is highly sensitive to the offsets and tilts in the system.

Different machine learning frameworks have been proposed in the area of the use of OAM EM modes for data transmission, most of those linked to free-space optical data transmission. Due to their distinct spatial distributions, the individual modes can be well recognized by machine learning tools. The OAM wave demodulation based on the *k*-nearest neighbors, naive Bayes classifier, and artificial neural networks

have been proposed; however, the results obtained by the use of convolutional neural networks (CNN) seem to be most promising. This could be a consequence of an inserted convolution-based feature extraction block and the enhancement of prediction accuracy by reducing the parametric complexity of the high-dimensional data. In this work, we will be using data from the millimeter-wave frequency range, generated assuming the UCA consisting of a limited number of antennas at both the transmitting and the receiving side. We describe the procedure of customizing, training, and testing a convolutional neural network (CNN), to perform the OAM mode recognition and to correct the offsets and tilts in a wireless communication system that uses OAM modes. A sufficient quantity of data for machine learning was generated using fast calculations of OAM EM fields based on the Hertz dipole method, as reported in (Ilić, Trajković, *et al*., 2023). We report on the success of training procedures when using different data sets as inputs (regarding the number of OAM EM field instances, and different considered scenarios). We put the obtained results into perspective by discussing our work and the results in the same field from the available literature. In conclusion, the CNN could be a useful tool in the mitigation of consequences of errors in positioning the transmitting and the receiving antenna arrays in an OAM-based system.

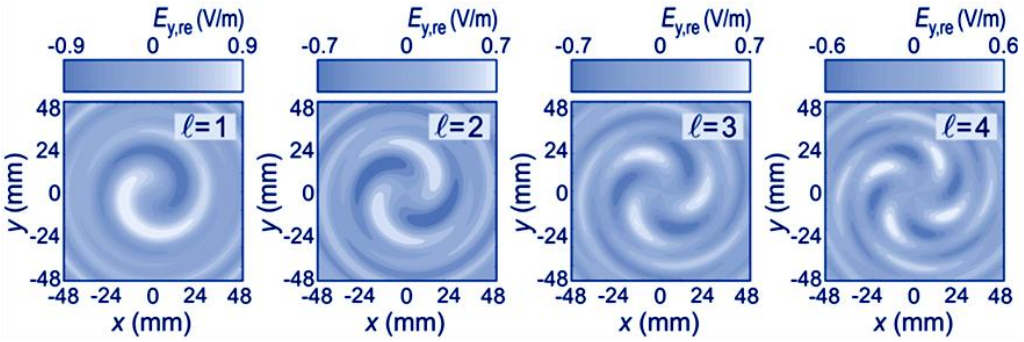


Fig. 1. Typical instantaneous EM fields of the first four OAM modes, represented by the real component of the electric field vector. To properly perform the stream separation procedure by the receiver, the transmitter and the receiver have to be well aligned with the system axis.

Acknowledgments: The authors acknowledge funding provided by the Institute of Physics Belgrade and School of Electrical Engineering, University of Belgrade, through the grants by the Ministry of Science, Technological Development, and Innovations of the Republic of Serbia.

Keywords: Machine learning (ML), convolutional neural networks (CNN), wireless communication systems (WCS), antenna arrays, orbital angular momentum (OAM) waves

References:

[1] Chen, S., Liang, Y., Sun, S., *et al*. (2020). Vision, requirements, and technology trend of 6G: How to tackle the challenges of system coverage, capacity, user data-rate and movement speed. *IEEE Wirel. Commun*., vol. 27(2), pp. 218–228.

[2] Ilić, A.Ž., Vojnović, N.M., Savić, S.V., *et al*. (2023). Optimized planar printed UCA configurations for OAM waves and the associated OAM mode content at the receiver. *Int. J. Commun. Syst*., vol. 36, e5623.

[3] Ilić, A.Ž., Trajković, J.Z., Savić, S.V., Ilić, M.M. (2023). Near-field formation of the UCA-based OAM EM fields and short-range EM power flux profiles. *J. Phys. A: Math. Theor*., vol. 56(25), p. 255701.

[4] Tariq, F., Khandaker, M.R.A., Wong, K.-K., *et al*. (2020). A speculative study on 6G. *IEEE Wirel. Commun*., vol. 27(4), pp. 118–125.

[5] Yin, X.-L., Sang, H.-Q., Cui, X.-Z., *et al*. (2017). Offset tolerance of an orbital angular momentum optical communication system with angular deflection. *Opt. Commun*., vol. 393, pp. 34–39.

Machine Learning-Based Profiling of Migraine with Aura: From Binary Detection to Multilevel Complexity Classification Using Structural MRI

Katarina Mitrović[1], Andrej M. Savić[2], Aleksandra Radojičić[3,4], Marko Daković[5], Igor Petrušić[5]

[1]Independent Researcher, Serbia, kacam.ftn@gmail.com
[2]Science and Research Centre, School of Electrical Engineering, University of Belgrade, Belgrade, Serbia
[3]Headache Center, Neurology Clinic, Clinical Center of Serbia, Belgrade, Serbia
[4]Faculty of Medicine, University of Belgrade, Belgrade, Serbia
[5]Laboratory for Advanced Analysis of Neuroimages, Faculty of Physical Chemistry, University of Belgrade, Belgrade, Serbia

Migraine with aura (MwA) is a disabling neurological disorder characterized by fully reversible neurological symptoms, such as visual, somatosensory, speech, motor, brainstem, and/or retinal, that can precede or follow the headache phase (Headache Classification Committee of the International Headache Society (IHS), 2018). Accurate detection and characterization of MwA subtypes and aura complexity remain clinically challenging. Machine learning (ML) offers a promising approach for extracting clinically relevant patterns from neuroimaging data and has already yielded significant insights in the study of MwA (Lee, 2025). Our prior work demonstrated that ML models trained on cortical morphometry could classify MwA vs. HCs and simple MwA vs. complex MwA with high accuracy. However, no prior study has explored Migraine Aura Complexity Score (MACS) as a categorical target for classification or directly compared regression and classification performance on this phenotyping task (Mitrović, 2023a; Mitrović, 2023b). The goal of this work is to explore the applicability of a previously developed framework by testing it on a novel task: predicting MACS as a categorical target variable, as a step toward developing a comprehensive multi-task ML approach for MwA profiling using morphometric features derived from structural MRI.

This study included patients with MwA selected from a prior neuroimaging cohort, with detailed inclusion and exclusion criteria described elsewhere (Mitrović, 2023b). The current study builds upon our previously established pipeline (Mitrović, 2023a). The same input feature space and preprocessing strategy were used, with the 340 features derived from structural MRI of cerebral cortex (34 cortical regions, 2 hemispheres, and 5 measures: cortical thickness, surface area, volume, Gaussian curvature, and folding index). Feature selection was performed using the Extremely Randomized Trees algorithm to reduce dimensionality and retain 40 most informative features. ML models were optimized and trained for this multi-class classification problem, and evaluated using leave-one-out cross-validation, ensuring methodological consistency with our prior work. This approach not only validates the robustness of the original model architecture but also demonstrates its flexibility in addressing clinically relevant, fine-grained categorizations within the MwA population. For the purposes of this study, MACS was converted from a continuous to a categorical variable, forming three groups:

- Group 1: MACS 0–1 (mild aura),
- Group 2: MACS >1–4.5 (moderate aura),
- Group 3: MACS >4.5 (complex aura).

Among the tested models, Linear Discriminant Analysis (LDA) achieved the highest accuracy of 85%, followed by Logistic Regression (LR) and Naive Bayes (NB), both with an accuracy of 75%. The performance of k-Nearest Neighbors (KNN, k=3) reached 70%, while Support Vector Machine (SVM) achieved 62.50%. The lowest classification accuracy was observed with the Classification and Regression Tree (CART) model, which yielded an accuracy of 60%. These results are presented in Figure 1.

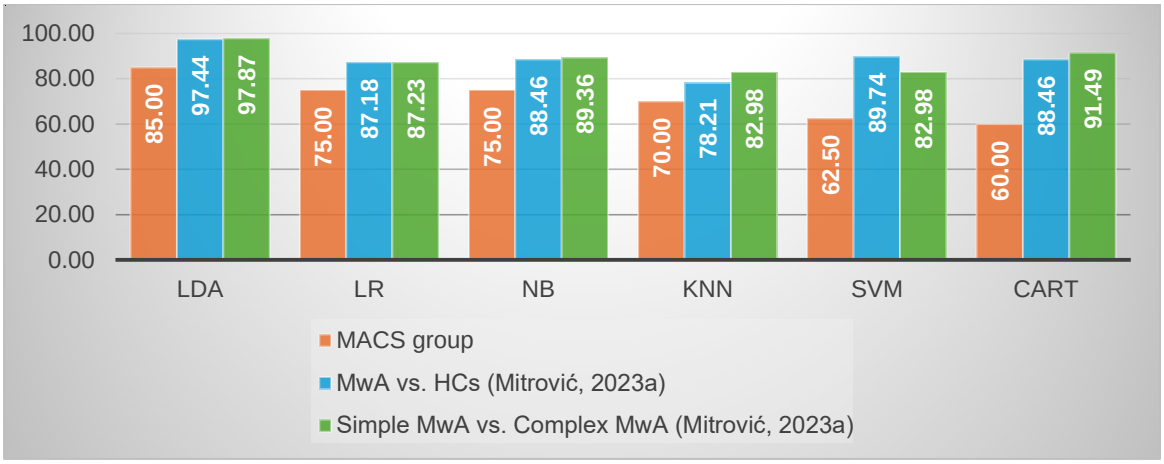


Fig. 1. MACS group classification results

The results of this study confirm that structural MRI features, when paired with robust ML techniques, can classify patients based on MACS group with satisfying accuracy. The high accuracy of the LDA-based model suggests that linear separation in a regularized low-dimensional subspace is well suited for this problem, likely due to the anatomical relevance of selected cortical features. The current classification task introduces a more nuanced, three-level output based on the MACS. This additional complexity likely contributes to the lower average model performance observed here, but also reflects a more granular and clinically meaningful stratification of the MwA population. Importantly, the reuse and adaptation of our previously validated pipeline, including the feature selection method and cross-validation strategy, allowed for direct methodological continuity and result comparability. While promising, this study is not without limitations. The relatively small sample size imposes constraints on the generalizability of findings and prevents more robust testing of deep or ensemble models. These findings support the feasibility of AI-driven phenotyping in migraine research and open avenues for future studies involving larger, more diverse datasets and multimodal inputs.

Keywords: Machine learning, LDA, Migraine with aura, MRI

References:

[1] Headache Classification Committee of the International Headache Society (IHS). (2018). The International Classification of Headache Disorders, 3rd edition. Cephalalgia, 38(1), 1–211. https://doi.org/10.1177/0333102417738202

[2] Lee, W., & Chu, M. K. (2025). The current role of artificial intelligence in the field of headache disorders, with a focus on migraine: A systematic review. Headache and Pain Research, 26(1), 48–65. https://doi.org/10.62087/hpr.2024.0024

[3] Mitrović, K., Petrušić, I., Radojičić, A., Daković, M., & Savić, A. (2023a). Migraine with aura detection and subtype classification using machine learning algorithms and morphometric magnetic resonance imaging data. Frontiers in Neurology, 14, 1106612. https://doi.org/10.3389/fneur.2023.1106612

[4] Mitrović, K., Savić, A. M., Radojičić, A., Daković, M., & Petrušić, I. (2023b). Machine learning approach for Migraine Aura Complexity Score prediction based on magnetic resonance imaging data. The journal of headache and pain, 24(1), 169. https://doi.org/10.1186/s10194-023-01704-z

Using Machine Learning to Predict Space Weather's Effect on Precipitation-Induced Floods from Solar Activity Time Series

Sonja D. Radenković[1], Milan Radovanović[2] & Slavica Malinović-Milićević[2]

[1]Belgrade Banking Academy – Faculty of Banking, Insurance and Finance, Union University, Zmaj Jovina 12, 11000 Belgrade, Serbia.

[2]Geographical Institute "Jovan Cvijic" SASA, Djure Jaksica 9, 11000 Belgrade, Serbia., sonja.radenkovic@bba.edu.rs, m.radovanovic@gi.sanu.ac.rs, s.malinovic-milicevic@gi.sanu.ac.rs

The purpose of this research is to identify the hidden dependencies that may exists between the flow of particles originated from the Sun and the floods that are caused by precipitation in the United Kingdom (UK). The investigation takes into account twenty floods that occurred between October 2001 and December 2019. The parameters of solar activity were utilized as input data for the model, while the data on precipitation that occurred 10 days prior to and during each flood event were used as output data for the model. The time lag of 0–9 days was taken into consideration during the research. The degree of randomness for the time series of input and output parameters was determined by the use of correlation analysis, which was carried out through the process. In order to establish a possible causal connection between the two variables, machine learning categorization predictive modeling was applied. We utilized two different methodologies: the decision tree and the random forest. From 0 to 9 days in advance, we conducted an analysis to see how accurate categorization models projected the future. Characteristics such as the Proton density with a time lag of nine, differential proton flux in the range of 310–580 keV, and ion temperature were shown to be the most critical factors for flood predicting. The research presented in this study has demonstrated that the decision tree model is superior in terms of accuracy and sufficiency when it comes to forecasting the occurrence of floods caused by precipitation up to nine days in advance, with an accuracy rate of 91%. By enhancing technical capabilities, utilizing enhanced machine learning approaches, and making use of big data sets, it is possible to increase the knowledge of the physical link between the solar wind and tropospheric weather, which in turn can assist improve severe weather forecasting. This was confirmed by the findings of this research.

Keywords: Machine Learning, Solar Activity, Precipitation, Floods, Classification, Modelling

References:

[1] Nitka, W.; Burnecki, K. Impact of Solar Activity on Precipitation in the United States. Phys. A Stat. Mech. its Appl. 2019, 527, 121387.

[2] Milovanović, B.; Radovanović, M. Connection between Solar Activity and Atmospheric Circulation in Period 1891-2004. J. Geogr. Inst. Jovan Cvijić. SASA 2009, 59, 35–48

[3] Malinović-Milićević, S., Radovanović, M. M., Radenković, S. D., Vyklyuk, Y., Milovanović, B., Milanović Pešić, A., Milenković, M., Popović, V., Petrović, M., Sydor, P., & Gajić, M. (2023). Application of Solar Activity Time Series in Machine Learning Predictive Modeling of Precipitation-Induced Floods. Mathematics, 11(4), 795. https://doi.org/10.3390/math11040795.

[4] Radovanović, M.M., Malinović-Milićević S., Radenkovic, S. , Milenković, M., Milovanović, B., Milanović Pešić, A., Popović, V. (2022). Influence of Space Weather on Precipitation Nouced Floods – Applying of Solar Activity Time Series in the Prediction of Precipitation-Induced Floods by Using the Machine Learning in Proceedings of РАЦИОНАЛЬНОЕ ПРИРОДОПОЛЬЗОВАНИЕ: ТРАДИЦИИ И ИННОВАЦИИ. 20-22 October 2022. Moscow. pp. 90-97.

[5] Todorović, N.; Vujović, D. Effect of Solar Activity on the Repetitiveness of Some Meteorological Phenomena. Adv. Sp. Res. 2014, 54, 2430–2440.

[6] Vyklyuk, Y.; Radovanović, M.; Milovanović, B.; Leko, T.; Milenković, M.; Milošević, Z.; Milanović Pešić, A.; Jakovljević, D. Hurricane Genesis Modelling Based on the Relationship between Solar Activity and Hurricanes. Nat. Hazards 2017, 85, 1043–1062

[7] Vyklyuk, Y.; Radovanović, M.M.; Stanojević, G.B.; Milovanović, B.; Leko, T.; Milenković, M.; Petrović, M.; Yamashkin, A.A.; Pešić, A.M.; Jakovljević, D.; et al. Hurricane Genesis

Modelling Based on the Relationship between Solar Activity and Hurricanes II. J. Atmos. Solar-Terrestrial Phys. 2018, 180, 159–164.

Modeling soil-to-plant transfer factors of natural radionuclides with Random Forest algorithm

Milena Živković[1], Dragana Krstić[1], Taha Yeseen Wais[2], Berivan F. Namq[3]

[1]Department of Physcis, Faculty of Science, University of Kragujevac,Serbia, milena.zivkovic@pmf.kg.ac.rs

[2]Chemical, Biological and Radiological Safety and Security Division, University of Mosul, Mosul, Iraq, taha.yaseen93@uomosul.edu.iq

[3]Department of Basic Science, College of Dentistry, University of Kirkuk, Kirkuk, Iraq, berivan-farooq@uokirkuk.edu.iq

The transfer of radionuclides from soil to plants represents an important link in the exposure chain of humans and animals to ionizing radiation. Traditionally, the transfer factor (TF) is used in assessments, defined as the ratio of radionuclide concentration in the plant to that in the soil. This study focuses on natural radionuclides (^{238}U, ^{232}Th, and ^{40}K), whose concentrations in plants depend on numerous factors: soil type, pH value, organic matter content, exchangeable potassium, plant species, and others (Kanmi et al, 2024; Djelic et al., 2016). To overcome the limitations of conventional models, we explore the application of the Random Forest algorithm for: predicting TF values, identifying the most influential parameters, and detecting anomalous values that deviate from typical patterns (Urso et al., 2023). The dataset used in this study is based on transfer factor values published in the IAEA Technical Reports Series No. 472, which provides standardized soil-to-plant radionuclide transfer parameters derived from a wide range of international studies and numerous scientific papers (IAEA, 2010; Djelic et al., 2016; Nisbet et al., 2000). The dataset includes a range of environmental and biological variables. While this study is not geographically constrained, the methodology can be adapted for regional datasets. Plant samples were categorized based on species and analyzed plant parts, including leaves, fruits, and roots. Soils were classified by type, such as sandy, clayey, and organic. In addition, physico-chemical properties such as soil pH, organic matter content, and exchangeable potassium concentration were recorded for each sample.

Table 1. Overview of key variables included in the dataset

Category	Variable	Description
Plant characteristics	Plant species	Taxonomic classification of plant samples
	Plant part	Leaf, root, fruit, whole plant, etc.
Soil characteristics	Soil type	Sandy, clayey, organic
	Soil pH	Measured in standard pH units
	Organic matter content	Percentage of dry soil mass
	Exchangeable potassium (K^+)	Measured in mval/100g dry soil
Radioactivity data	Radionuclide concentration (soil)	Bq/kg dry mass, for ^{238}U, ^{232}Th, ^{40}K
	Radionuclide concentration (plant)	Bq/kg dry mass, for ^{238}U, ^{232}Th, ^{40}K
Derived variable	Transfer factor (TF)	C_plant / C_soil
	Log-transformed TF	ln(TF)

TF was calculated using the ratio of radionuclide concentration in plant tissue to that in the corresponding soil sample, expressed in Bq/kg of dry mass. Prior to analysis, the dataset was

cleaned to remove incomplete or inconsistent entries. TF values were log-transformed (log-TF) to reduce skewness, and all numerical features were standardized. The complete dataset was randomly split into a training set (80%) and a testing set (20%) for model development and evaluation.

Model development was performed using the *scikit-learn* library (or alternatively the *randomForest* package in R, depending on the programming environment). Hyperparameter tuning was conducted using k-fold cross-validation with $k = 5$, ensuring robust performance estimation and reduction of overfitting. Model evaluation was based on several commonly used metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), which collectively provide insight into both the accuracy and explanatory power of the regression model. Anomalous values were defined as those deviating by more than two standard deviations from the model's predicted values or exhibiting unusually high residuals in the test dataset. This approach enabled the identification of outliers that may suggest local environmental contamination, measurement errors, or unique soil-plant interactions. The Random Forest model demonstrated high predictive accuracy, achieving RMSE = 0.45, MAE = 0.35, and $R^2 = 0.86$. Among the most influential predictors were plant type, soil pH, exchangeable potassium, and organic matter content, highlighting the interplay between biological and physicochemical factors in radionuclide uptake. Several anomalous TF values were detected at sites characterized by a specific combination of low soil pH and elevated ^{238}U concentrations, indicating potential environmental irregularities.

Overall, the RF model outperformed traditional empirical and semi-mechanistic approaches in both accuracy and its ability to explain data variability. Categorical predictors, especially plant-related variables, accounted for a substantial portion of TF variation, emphasizing the importance of plant-specific traits in radioecological modeling. Detected anomalies suggest the need for further field investigation, as they may reflect either natural geochemical peculiarities or anthropogenic contamination.

Keywords: Soil-to-plant transfer, natural radionuclides, Random Forest, anomaly detection

References:

[1] Kanmi, A. S., Ibrahim, U., Goki, N. G., Rilwan, U., Sayyed, M. I., Wais, T. Y., Namaq, B. F., & Najam, L. A. (2024). Estimation of soil-to-plant transfer factor across six local government areas of Kwara state, Nigeria. *Journal of Environmental Radioactivity, 280*, 107548. https://doi.org/10.1016/j.jenvrad.2024.107548

[2] Djelic, G., Krstic, D., Stajic, J. M., Milenkovic, B., Topuzovic, M., Nikezic, D., Vucic, D., Zeremski, T., Stankovic, M., & Kostic, D. (2016). Transfer factors of natural radionuclides and ^{137}Cs from soil to plants used in traditional medicine in central Serbia. *Journal of Environmental Radioactivity, 158–159*, 81–88. https://doi.org/10.1016/j.jenvrad.2016.03.028

[3] Urso, L., Petermann, E., Gnädinger, F., & Hartmann, P. (2023). Use of random forest algorithm for predictive modelling of transfer factor soil-plant for radiocaesium: A feasibility study. *Journal of Environmental Radioactivity, 270*, 107309. https://doi.org/10.1016/j.jenvrad.2023.107309

[4] International Atomic Energy Agency. (2010). *Handbook of parameter values for the prediction of radionuclide transfer in terrestrial and freshwater environments* (Technical Reports Series No. 472). IAEA. https://www-pub.iaea.org/MTCD/Publications/PDF/trs472_web.pdf

[5] Nisbet, A. F., & Woodman, R. F. (2000). Soil-to-plant transfer factors for radiocesium and radiostrontium in agricultural systems. *Health Physics, 78*(3), 279–288. https://doi.org/10.1097/00004032-200003000-00005

NATURAL LANGUAGE PROCESSING

Multimodal Intelligence for Medical Text Extraction: Benchmarking Vision LLMs Against Classic Pipelines in Serbian

Aldina Avdić[1], Adela Ljajić[2], Ulfeta Marovac[1]

[1]Department of Technical-Technological Sciences, State University of Novi Pazar, Vuka Karadžića bb, Serbia, [apljaskovic, umarovac]@np.ac.rs

[2]The Institute for Artificial Intelligence, Research and Development of Serbia, Fruškogorska 1 Serbia, adela.ljajic@ivi.ac.rs

In today's digital environment, medical reports are increasingly being shared on social media platforms, often without adequate protection of personal information. This paper presents a comparative analysis of two approaches for recognizing personal names in images of medical reports written in Serbian: a traditional rule-based method and a modern approach based on Large Language Models (LLaMA), optimized using the QLoRA technique.

For this study, a dataset of 71 images of medical reports was collected from a closed Facebook group dedicated to insulin resistance patients [Avdić & Marovac, 2022]. These images contain identifiable personal information such as names and surnames, offering a realistic scenario for evaluating de-identification systems. An additional evaluation was conducted on an independent dataset collected from various online platforms to assess the robustness of both methods.

The rule-based approach uses the Tesseract OCR engine to extract text from images, followed by handcrafted rules and regular expressions to identify personal data. If this step fails, the extracted Serbian text is automatically translated into English using the Python translate library[4]. Then, the SpaCy NER tool, trained on English data, is applied to detect entities of type PERSON. This method is easy to implement and requires no training, but it relies heavily on OCR quality of machine translation accuracy [Mathew et al., 2021, Hirota et al., 2021].

The LLM-based approach utilizes a vision-language architecture based on a fine-tuned LLaMA model with QLoRA-based quantized adapters, enabling efficient inference on limited hardware [Demirhan & Zadrozny, 2024]. Unlike the rule-based method, which relies on OCR and text translation, this approach processes the original image of the medical report directly, extracting both textual and visual cues. It identifies personal names by jointly interpreting layout, typography, and sentence-level context. Its performance was evaluated under varying image quality conditions to simulate real-world scenarios.

Experimental results indicate that the LLM-based approach provides superior precision and recall, especially when dealing with noisy or low-quality text. In contrast, the rule-based method is faster to deploy and requires fewer resources, making it suitable for resource-constrained environments. A hybrid approach [de Gast, 2022], combining both methods, appears to be the most effective path toward building a robust system for de-identifying medical image data.

These approaches contribute to stronger patient privacy protection and enable secure use of medical records for research, further analysis, and AI model training.

Keywords: De-identification, Medical Report Images, Named Entity Recognition (NER), Large Language Models (LLM), Serbian Language Processing

References:

[4] https://pypi.org/project/translate/

[1] Avdić, A., & Marovac, U. (2022). *Prepoznavanje imena na slikama lekarskih izveštaja na srpskom jeziku u cilju zaštite ličnih podataka*. Zbornik radova, LXVI Konferencija ETRAN, Novi Pazar, 6–9. jun 2022.

[2] de GAST, K. A. I. (2022). *Medical VQA analysis and improvement using clinical knowledge graphs* [Master's thesis, Politecnico di Milano]. https://www.politesi.polimi.it/handle/10589/188320

[3] Demirhan, H., & Zadrozny, W. (2024). *Survey of Multimodal Medical Question Answering*. MedAI, 4(1), 4. https://www.mdpi.com/2673-7426/4/1/4

[4] Hirota, Y., Sugawara, S., & Aizawa, A. (2021). *Visual Question Answering with Textual Representations for Document Images*. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). https://openaccess.thecvf.com/content/ICCV2021/papers/Hirota_Visual_Question_Answering_With_Textual_Representations_for_Document_Images_ICCV_2021_paper.pdf

[5] Mathew, M., Iyyer, M., & Jawahar, C. V. (2021). *DocVQA: A Dataset for VQA on Document Images*. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). https://openaccess.thecvf.com/content/WACV2021/papers/Mathew_DocVQA_A_Dataset_for_VQA_on_Document_Images_WACV_2021_paper.pdf

Using a generative LLM to create an annotated dataset in Serbian: Outcomes and lessons learned

Gorana Gojić[1], Milena Oparnica[1], Darija Medvecki[1], Bojana Bašaragin[1]

[1]The Institute for Artificial Intelligence Research and Development of Serbia, Fruškogorska 1, Novi Sad, Serbia,
{gorana.gojic,milena.oparnica,darija.medvecki,bojana.basaragin}@ivi.ac.rs

In this talk, we present our work on generating a synthetic, annotated general-domain dataset designed to support the evaluation of IVIAN — an anonymization software tailored for the Serbian language (*IVIAN - Anonimizator Za Srpski Jezik*, 2024). This dataset is the first of its kind in terms of its fully synthetic nature and content, offering a novel contribution to the field of natural language processing (NLP) of Serbian.

With the rapid development of large language models (LLMs), there has been a growing interest in leveraging them to generate synthetic datasets to train and evaluate NLP solutions. This approach is particularly valuable in the context of low-resource languages, where the availability of human-annotated data is limited or nonexistent (Cvetanović & Tadić, 2024; Li et al., 2023). Synthetic datasets offer a scalable, privacy-protecting, and cost-effective alternative, and they have already been successfully used to train models for low-subjectivity NLP tasks (e.g. news topic classification, spam email detection), achieving performance comparable to models trained on real-world data (Li et al., 2023). Motivated by these advantages and the scarcity of available datasets, most of which focus only on a limited set of entity types, we decided to experiment with a synthetic dataset to evaluate our anonymization tool in the general domain.

The dataset comprises 300 emails with an equal split between Cyrillic and Latin scripts, spanning four distinct types of correspondence: private-to-private, private-to-business, business-to-business, and business-to-private. This email format was deliberately chosen as it allows for a rich and diverse presence of entity types relevant to real-world communication. Each email is annotated in the CoNLL-2002 format, following the BIO2 (IOB2) tagging scheme. The dataset covers 13 distinct types of entities – their distribution, along with the dataset itself, can be found on Hugging face[5].

The email content and entity annotations were generated using GPT-4o (OpenAI et al., 2024), with a knowledge cut-off date of October 2023, an input context length of 128k tokens, and a maximum output size of 4096 tokens. Content creation is performed through guided prompt design, which includes a task description, detailed instructions, formatting constraints, error minimization strategies, and few-shot examples illustrating the expected output. For the email generation task, we designed four distinct prompts, one for each correspondence type, allowing fine-grained control over the entity types generated for each correspondence type. We applied the same content creation approach when generating the annotations while adding real-world examples of manually annotated entities. Content generation is followed by a manual postprocessing pipeline to ensure consistency and quality of annotations.

This talk has three goals:

1. To provide a reproducible overview of our dataset creation process, including detailed recommendations for prompt engineering and postprocessing techniques, and discuss the challenges.

[5] https://huggingface.co/datasets/goranagojic/SerbianEmailsNER

2. To introduce and discuss strategies for increasing the lexical and numerical variability of entities within the generated texts, which is crucial for robust evaluation of anonymization tools.

3. To present and announce the release of the evaluation dataset, offering a general-domain benchmark with diverse entities for researchers and developers working on the tasks of named entity recognition and de-identification.

By sharing both the methodology and the resulting dataset, we aim to encourage further research into synthetic data generation.

Keywords: Natural language processing, generative artificial intelligence, synthetic dataset, named entity recognition

References:

[1] Cvetanović, A., & Tadić, P. (2024). *Synthetic Dataset Creation and Fine-Tuning of Transformer Models for Question Answering in Serbian* (arXiv:2404.08617). arXiv. https://doi.org/10.48550/arXiv.2404.08617

[2] IVIAN - *Anonimizator za srpski jezik*. (2024). https://hci.institutonline.ai/anonimizator

[3] Li, Z., Zhu, H., Lu, Z., & Yin, M. (2023). *Synthetic Data Generation with Large Language Models for Text Classification: Potential and Limitations* (arXiv:2310.07849). arXiv. https://doi.org/10.48550/arXiv.2310.07849

[4] OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774

Towards Linguistic Completeness in Knowledge Graphs: Generating Serbian Inflections with Language Models

Milica Ikonić Nešić[1], Saša Petalinkar[2], Ranka Stanković[3]

University of Belgrade, Faculty of Philology, Serbia, milica.ikonic.nesic@fil.bg.ac.rs
University of Belgrade, Serbia, sasa5linkar@gmail.com
University of Belgrade, Faculty of Mining and Geology, Serbia, ranka@rgf.rs

Named entity recognition (NER) and named entity linking (NEL) to open knowledge bases such as Wikidata remain particularly challenging for low-resource languages like Serbian (Ikonić Nešić et al., 2024). Although Wikidata offers a vast repository of entities, most aliases are available only in their normalized (lemma) form. This significantly limits their applicability in downstream tasks such as NER and NEL, where inflected variants are frequent in real-world textual data. While the issue of missing items in Wikidata has been successfully addressed in previous work (Ikonić Nešić et al., 2022), the lack of morphological coverage, especially for languages with rich inflectional systems, still limits the effectiveness of entity recognition and linking.

The lexical database Leximirka (Lazić and Škorić 2019, Stankovic et. al. 2018) presents one of the central resources for processing Serbian named entities. Beyond offering rich semantic markers that facilitated the targeted retrieval of specific named entity categories from the corpus, a critical feature of Leximirka for this research was its comprehensive storage of the complete inflectional paradigm for each lemma. This capability proved invaluable: following the extraction of QIDs from Wikidata and their successful linking to the corresponding lexical units within Leximirka, the availability of all inflected forms enabled the precise retrieval of the correct grammatical forms for all grammatical cases of the named entities linked to Wikidata. Leveraging Leximirka's detailed inflectional data was key to effectively handling the complex morphology inherent in Serbian proper names during this research.

To address this gap, this research proposes an exploratory, proof-of-concept approach that leverages large language models (LLMs), specifically, gpt-5 by Open Ai, to automatically generate Serbian inflected forms for Wikidata entities. For each Wikidata item, prompts based on the entity's name and short description are constructed, providing the language model with sufficient context to infer the entity type (e.g., person, location, organization) and likely inflectional behavior. The LLM is then tasked with generating all plausible case, number, and gender variants for that entity, simulating the work of a skilled human annotator or lexicographer.

The quality and coverage of the generated inflected forms are systematically evaluated against the Leximirka database. Precision, recall, and coverage of the LLM-generated forms are measured through this comparison, recurring error types (e.g., ambiguous gender, irregular declension) are identified, and the practical impact for NER/NEL pipeline performance is assessed. Qualitative findings are taken into account, including the LLM's ability to handle edge cases such as foreign names, acronyms, and multi-word entities.

By automating the generation of inflected forms, a scalable alternative to manual annotation is provided, thereby enhancing the practical utility of Wikidata and similar resources for downstream NLP tasks in Serbian. Systematic gaps are highlighted by our error analysis, and directions for future improvements are suggested, such as integrating explicit context or combining neural and rule-based methods. A hybrid approach is indicated by the findings: inflected variants for large-scale resources can be bootstrapped by LLMs, while ambiguous or error-prone cases can be handled by manual or rule-based post-processing.

It has been shown by previous work (Petalinkar, 2024) that LLM can be successfully leveraged to enrich Serbian NLP resources through prompt engineering and synthetic data generation. Although sentiment lexicons were the focus of that work, the underlying methodology, using LLMs to generate linguistically relevant variants, can be adapted to the challenge of producing comprehensive inflectional forms for named entities, as is proposed here.

The implications of this work are seen beyond Serbian and Wikidata. Similar constraints are faced by many morphologically rich, low-resource languages, and the approach outlined here can be readily adapted to other contexts, including digital lexicography, corpus annotation, and machine translation. Furthermore, current trends in multilingual NLP are aligned with the use of LLMs for resource enrichment, where transformer-based models are increasingly leveraged to close resource gaps, improve coverage, and accelerate the development of robust tools for underrepresented languages. The feasibility and expectation of using LLMs for morphological enrichment of knowledge base entities are demonstrated by this study, with a focus on Serbian as a case study. Future efforts to integrate LLM-based morphology generation into broader entity linking workflows for highly inflected languages are informed by the results and analysis presented in this research. A foundation for further enriching the Wikidata knowledge base with new properties, specifically those directly related to case inflection across different languages, is provided.

Acknowledgments: This work was supported by the Science Fund of the Republic of Serbia #7276, Text Embeddings – Serbian Language Applications, TESLA.

Keywords: Named Entity Linking, language models, Wikidata, Leximirka

References:

[1] Ikonić Nešić, M., Petalinkar, S., Stanković, R., Utvić, M., & Kitanović, O. (2024, September). SrpCNNeL: Serbian model for named entity linking. In *2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS)* (pp. 465–473). IEEE.

[2] Ikonić Nešić, M., Stanković, R., Schöch, C., & Škorić, M. (2022). From ELTeC text collection metadata and named entities to linked-data (and back). In *Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference* (pp. 7–16). European Language Resources Association (ELRA).

[3] Stanković, R., Krstev, C., Lazić, B., & Škorić, M. (2018). Electronic dictionaries – From file system to lemon-based lexical database. In J. P. McCrae, C. Chiarcos, T. Declerck, J. Gracia, & B. Klimek (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Paris, France: European Language Resources Association (ELRA).

[4] Lazić, B., & Škorić, M. (2019). From DELA based dictionary to Leximirka lexical database. *Infotheca – Journal for Digital Humanities, 19*(2), 81–98. https://doi.org/10.18485/infotheca.2019.19.2.4

[5] Petalinkar, Saša. 2024. "Creating a Synthetic Evaluation Dataset for Serbian SentiWordNet Using Large Language Models." *Infotheca* 24 (1). https://doi.org/10.18485/infotheca.2024.24.1.3.

Fake News Detection in Serbian using Large Language Models and Knowledge Graphs

Danka Jokić[1], Ranka Stanković [2], Ana Kovačević[3]

[1]University of Belgrade, Serbia, danka.jokic@afrodita.rcub.bg.ac.rs
[2]Faculty of Mining and Geology, University of Belgrade, Serbia, ranka.stankovic@rgf.bg.ac.rs
[3]Faculty of Security Studies, University of Belgrade, Serbia, kana@fb.bg.ac.rs

The proliferation of fake news and misinformation poses a significant global challenge with profound societal implications (Kumar et al., 2025). As these incidents become more frequent, the necessity for automated detection systems intensifies. Fake news detection, a natural language processing task, involves identifying and classifying news articles or social media messages as real or fake. The goal is to develop algorithms that can automatically detect and flag fake news, thereby mitigating misinformation and promoting the dissemination of accurate information. While substantial resources exist for the English language, developing detection systems for other languages, particularly low-resource languages such as Serbian, remains challenging due to the limited availability of annotated datasets, fact-checking resources, and computational models (Sivanaiah et al., 2022).

Large Language Models (LLMs) have shown solid performance in detecting fake news (Sellami et al., 2024). However, they often lack explicit factual grounding and are prone to hallucinations, especially in languages with sparse training data. In contrast, Knowledge Graphs (KGs) provide structured factual representations, offering a reliable external knowledge base that can integrated into LLM-based fake news detection pipelines to enhance contextual representations and improve the accuracy of fake news detection. Knowledge-enhanced Language Models for Fake News Detection integrate structured factual knowledge from one or multiple sources into pre-trained LLMs (Whitehouse et al., 2022). Knowledge graphs are either external or specially designed internal knowledge graphs based on available data in the dataset and obtained by extracting subject-predicate-object relationships from claims and justifications to compute factual alignment (Koloski et al., 2022; Kumar et al., 2025). There is no recent work that explores either the usage of LLMs or the synergy of KGs and LLMs for fake news detection in the Serbian language.

To address these gaps, this research aims to explore the synergy between LLMs and KGs to enhance fake news detection in the Serbian language. We will present the framework that we will use for the automatic detection of fake news in Serbian, which could be expanded to a multilingual context. Additionally, we will introduce our work on collecting data for a Serbian-language fake news dataset from fact-checking portals and our initial experiments with applying machine learning and generative AI for detecting fake news. Moreover, we will report our experiments with real and AI-generated fake news. This work contributes to the growing field of trustworthy AI and misinformation resilience, with applications in journalism, cybersecurity, and social media regulation.

Keywords: fake news, LLMs, knowledge graph

References:

[1] Koloski, B., Perdih, T. S., Robnik-Šikonja, M., Pollak, S., & Škrlj, B. (2022). Knowledge graph informed fake news classification via heterogeneous representation ensembles. Neurocomputing, 496, 208-226.
[2] Kumar, A., Kumar, P., Yadav, A., Ahlawat, S., & Prasad, Y. (2025, January). KGFakeNet: A Knowledge Graph-Enhanced Model for Fake News Detection. In Proceedings of the Workshop on Generative AI and Knowledge Graphs (GenAIK) (pp. 109-122).
[3] Sallami, D., Chang, Y. C., & Aïmeur, E. (2024). From deception to detection: The dual roles of large language models in fake news. arXiv preprint arXiv:2409.17416.
[4] Sivanaiah, R., Ramanathan, N., Hameed, S., Rajagopalan, R., Suseelan, A. D., & Thanagathai, M. T. N. (2022, November). Fake news detection in low-resource languages. In International

Conference on Speech and Language Technologies for Low-resource Languages (pp. 324-331). Cham: Springer International Publishing.
[5] Whitehouse, C., Weyde, T., Madhyastha, P., & Komninos, N. (2022, May). Evaluation of fake news detection with knowledge-enhanced language models. In Proceedings of the international AAAI conference on web and social media (Vol. 16, pp. 1425-1429).

Benchmarking Machine-Generated Text Detection in Serbian

***Marko Koprivica*[1]**

[1]Faculty of Organizational Sciences, Department of Software Engineering, University of Belgrade, Jove Ilica 154, Belgrade, Serbia, marko.koprivica@hotmail.com

Large Language Models (LLMs) have profoundly impacted a wide range of sectors, reshaping how individuals interact with information, automate tasks, and process natural language. Even though the conceptual roots of LLMs can be traced back to the early development of natural language processing (NLP) (Maatouk et al., 2024), the release of Chat Generative Pre-Trained Transformer (ChatGPT) in November 2022 denoted a pivotal moment in the evolution of the field (Korseberg & Elken, 2024). However, misuse has raised serious concerns—particularly in educational settings, where students may use LLMs to generate essays, assignments, or problem solutions (Liu et al., 2023). This has led to growing interest in the development of machine-generated text detection methods, aimed at distinguishing AI-generated content from human-written text to prevent abuses such as academic misconduct, spam, and disinformation (Dou et al., 2020).

The growing number of machine-generated text detection methods proposed since 2023 is encouraging. However, the absence of standardized benchmark datasets, the use of diverse evaluation metrics, and the limited reproducibility of existing approaches pose significant challenges to reliably comparing their effectiveness. While many methods are evaluated on custom datasets consisting of human and machine-generated texts, several benchmark datasets do exist across various languages and domains [Wu et al., 2025]. Nevertheless, no such benchmark has yet been developed for texts in the Serbian language.

In this work, we introduce a benchmark dataset specifically designed for Serbian, consisting of authentic news texts and their counterparts generated by large language models (LLMs). The dataset is accompanied by evaluation using standard metrics, including accuracy, precision, recall, F1-score, and AUC. Machine-generated texts (MGTs) were produced using three prominent large language models: GPT-5, Claude, and DeepSeek. To ensure consistency and realism, the generation process was prompt-driven. For each human-written text (HWT), we set the target word count N to match the original text length, rounded to the nearest hundred. The prompt provided to the models was structured as follows: "*Write a news article in N words with the following headline: <headline>.*" To assess model robustness, we further consider three types of adversarial attacks: paraphrasing, which rephrases the original text while preserving its meaning; random spacing, which disrupts typical word and sentence spacing to confuse tokenization and text structure; and adversarial perturbations, which introduce subtle character-level modifications (e.g., letter swaps or symbol insertions) designed to evade detection without significantly altering the text's readability or semantics.

The dataset was evaluated using detection tools such as GPTZero and DetectGPT, which serve as representative examples of current approaches to distinguishing machine-generated from human-authored texts. The findings highlight notable differences in detection performance: some tools exhibited greater sensitivity to outputs from particular language models, while others were less capable of reliably identifying machine-generated content. Moreover, the analysis suggests that advanced models increasingly produce text that is classified as human-written, whereas the classification of genuine human texts often remained inconclusive or inconsistent across detectors. For instance, DetectGPT consistently struggled to classify any content as machine-generated, while GPTZero achieved higher success rates in identifying outputs from models such as Cloud and DeepSeek. At the same time, GPTZero frequently produced uncertain classifications for genuine human texts, and advanced models such as GPT-5 were predominantly misclassified as human-written.

Keywords: Large language models (LLMs), generative AI, detection, LLM-generated text, benchmarking

References:

[1] Dou, Y., Ma, G., Yu, P., Xie, S., 2020. Robust Spammer Detection by Nash Reinforcement Learning. https://doi.org/10.1145/3394486.3403135

[2] Korseberg, L., Elken, M., 2025. Waiting for the revolution: how higher education institutions initially responded to ChatGPT. High Educ 89, 953–968. https://doi.org/10.1007/s10734-024-01256-4

[3] Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., Lyons, K., Gašević, D., Chen, G., 2023. Can large language models write reflectively. Computers and Education: Artificial Intelligence 4, 100140. https://doi.org/10.1016/j.caeai.2023.100140

[4] Maatouk, A., Piovesan, N., Ayed, F., De Domenico, A., Debbah, M., 2025. Large Language Models for Telecom: Forthcoming Impact on the Industry. IEEE Communications Magazine 63, 62–68. https://doi.org/10.1109/MCOM.001.2300473

[5] Wu, J., Yang, S., Zhan, R., Yuan, Y., Chao, L.S., Wong, D.F., 2025. A Survey on LLM-Generated Text Detection: Necessity, Methods, and Future Directions. Computational Linguistics 51, 275–338. https://doi.org/10.1162/coli_a_00549

Analysis of LLMs and Classical NLP Systems for Investment Sentiment

***Petar Lakčević*[1]**

Faculty of Organizational Sciences, University of Belgrade, Serbia,
petar.lakcevic@fon.bg.ac.rs

This paper researches the application of large language models (LLM) for sentiment analysis in the context of financial investing. Specifically, it focuses on sentiment analysis of news articles and forum posts related to stock trading. The primary objective was to evaluate how efficiently and effectively various LLMs can identify sentiment in real-world financial texts which often exhibit ambiguity, mixed emotional signals, and often other overlapping sentiments. This research emphasizes the ability of LLMs to recognize investment sentiment, that is whether the overall tone of the message suggest a positive or negative outlook from an investor's perspective, rather than general emotional sentiment, which may be misleading in financial interpretation.

This evaluation should be seen as a general, exploratory test, since the selected LLMs differ significantly in architecture, pricing model, and intended use cases, and are positioned differently in the current market landscape (Leon, 2024).

Three major language models were used in the research: gpt-3.5-turbo, gpt-4o, and DeepSeek-7B, while Stanford CoreNLP was applied as a control mechanism (a classic NLP keyword-based system). All models were tested on an identical set of 500 manually labeled financial news articles gathered from a few chosen news websites and forums, with each text containing an explicitly labeled sentiment on a five-point scale: very negative, negative, neutral, positive, or very positive. All LLMs were given the same predefined prompts, to ensure consistency in the query and allow for reliable comparisons between model.

Each model was applied individually to the same dataset. The outputs obtained from the LLM models were normalized to match the format of the target sentiment categories, to allow for direct comparison with manually labeled values. Evaluation metrics included classification accuracy, average processing time per document, and cost (in terms of API calls if applicable).

The dataset itself posed a particular challenge due to the complexity of many news stories, which often feature multiple and conflicting sentiment messages, making it difficult even for humans to clearly determine the dominant sentiment from an investor perspective. In addition, the texts were manually sorted by sentiment recognition difficulty, so the dataset was divided into several groups: 200 news stories with the clearest sentiment, 200 with moderately complex, and 100 with the most ambiguous expressions, which allowed for further analysis of the model's performance in relation to the task difficulty.

The best results in terms of accuracy were shown by the gpt-4o, which was expected given that it is the model with the strongest reasoning capabilities in this comparison (Singgalen, 2024). It also performed well on complex texts with multiple sentiments and subtle nuances.

The gpt-3.5-turbo offered the best price-performance ratio, but its accuracy declined somewhat as the texts became more complex and sentiment more hidden. This behavior was expected since gpt-3.5-turbo model is primarily optimized for efficiency and broad accessibility rather than deep contextual reasoning (Yang et al., 2024).

DeepSeek-7B, as an open model, showed solid results, but had significantly higher latency due to local inference.

Stanford CoreNLP was by far the fastest and cheapest, and proved surprisingly adequate in cases of simple, clearly expressed sentiments. In the group of the easiest 200 texts, all models gave approximately the same results, which indicates that in such cases it is most cost-effective to use classic NLP tools, as they require minimal resources and have the lowest amortization cost. However, as text weight increases, LLMs (especially gpt-4o) increasingly dominate in prediction accuracy and stability, making them more suitable for complex or ambiguous content.

Future work will focus on expanding the dataset across a longer time horizon, additional input sources, broader range of financial subjects as well as expanding the number of LLMs tested for this purpose. It will also be important to test multiple models within the same class (e.g., comparing GPT-4o with Claude 3.5 or similar) to better understand the relative strengths of different architectures under the same conditions. Further directions could also include training a domain specific financial sentiment model with a focus on real time sentiment analysis.

Keywords: Large Language Models, Investment Sentiment, Financial Text Analysis, Sentiment Classification

References:

[1] Leon, M. (2024). Comparing LLMs using a Unified Performance Ranking System. International Journal of Artificial Intelligence & Applications, 15(4), 33–46. https://doi.org/10.5121/ijaia.2024.15403

[2] Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., & Hu, X. (2024). Harnessing the power of LLMs in Practice: A survey on ChatGPT and Beyond. ACM Transactions on Knowledge Discovery From Data, 18(6), 1-32. https://doi.org/10.1145/3649506

[3] Singgalen, Y. A. (2024). Analyzing an Interest in GPT 4o through Sentiment Analysis using CRISP-DM. Journal of Information Systems and Informatics, 6(2), 882–898. https://doi.org/10.51519/journalisi.v6i2.740

Efficient Fine-Tuning of LLaMA 3.2 Vision for Serbian Language Generation

Adela Ljajić[1], Marko Zubac[1], Miloš Košprdić[1], Nikola Milošević[1], Dragiša Mišković[1]

The Institute for Artificial Intelligence Research and Development of Serbia, Fruškogorska 1, Novi Sad, Serbia, {adela.ljajic, marko.zubac, miloš.kosprdic, nikola.milosevic, dragisa.miskovic}@ivi.ac.rs

Recent advancements in large language models (LLMs) have primarily centered on high-resource languages, leaving low-resource languages such as Serbian significantly underrepresented. In this work, we present the methodology and results of fine-tuning the LLaMA 3.2 Vision 11B model for causal language modeling (CLM) on a corpus of Serbian text. We employed Parameter-Efficient Fine-Tuning (PEFT) using QLoRA (Dettmers et al., 2023) with 4-bit quantization, making the process computationally feasible on a single NVIDIA A100 40GB GPU. The goal was to adapt a state-of-the-art large language model to better understand and generate Serbian text.

We employed LoRA adapters (Hu et al., 2021) in conjunction with BitsAndBytes 4-bit quantization (NF4), using the Hugging Face transformers and peft libraries. The adapters targeted only the attention mechanism, specifically the query and value projection layers (q_proj, v_proj), with a rank of 16. Training was conducted over 2 epochs with a context window of 1024 tokens, batch size of 4, and gradient accumulation of 8 steps, yielding an effective batch size of 32.

The dataset consisted of around 30,000 samples drawn from the SrWaC corpus (Ljubešić & Klubička, 2014), chunked to 1024 tokens per sample. Validation was performed using perplexity and token-level loss metrics. Although the model supports multimodal inputs, we used it in text-only mode, leveraging its strong language modeling capabilities for CLM.

The training dynamics show a significant reduction in loss over epochs, confirming the model's capacity to internalize Serbian syntax and vocabulary. We evaluated the fine-tuned adapter using the perplexity metric on a held-out portion of the training dataset. We calculated average perplexity using greedy decoding (temperature = 0) to ensure stable and reproducible results by eliminating sampling randomness. Lower perplexity indicates the model's better confidence in predicting the next token.

Our adapter-tuned model achieved significantly lower perplexity than both LLaMA 3.2 Vision 11B (base) and LLaMA 3.2 Vision 11B Instruct, demonstrating stronger adaptation to Serbian linguistic patterns.

Our results highlight the feasibility of adapting large-scale base models to low-resource languages using resource-efficient techniques. These findings open new possibilities for scalable LLM adaptation in constrained settings and emphasize the importance of language-specific training to ensure inclusivity in AI development. Moreover, our results demonstrate that 4-bit QLoRA fine-tuning of a vision-capable LLM can effectively adapt to monolingual text generation tasks in low-resource languages and pave the way for future multimodal extensions. A notable limitation of this approach is the long training time; in our case, the process lasted 20 days on a single NVIDIA A100 40GB GPU when fine-tuning on the full SrWaC corpus, which may pose challenges for time-constrained development.

In future work, we plan to extend this approach by fine-tuning the adapter on a larger dataset and applying the model’s vision adapter to enable multimodal understanding of Serbian content in images, such as scanned documents or social media posts.

Keywords: Causal language modeling, Fine-tuning, Serbian language, LLaMA, QLoRA.

References:

[1] Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). *QLoRA: Efficient Finetuning of Quantized LLMs* (No. arXiv:2305.14314). arXiv. https://doi.org/10.48550/arXiv.2305.14314

[2] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). *LoRA: Low-Rank Adaptation of Large Language Models* (No. arXiv:2106.09685). arXiv. https://doi.org/10.48550/arXiv.2106.09685

[3] Ljubešić, N., & Klubička, F. (2014). {bs,hr,sr}WaC - Web Corpora of Bosnian, Croatian and Serbian. In F. Bildhauer & R. Schäfer (Eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)* (pp. 29–35). Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-0405

Text validation in NLP applications: A Chain of Responsibility approach

Dragica Ljubisavljević[1], Ana Korunović[2]

[1]Faculty of Organizational Sciences, University of Belgrade, Serbia,
dragica.ljubisavljevic@fon.bg.ac.rs
[2]Faculty of Organizational Sciences, University of Belgrade, Serbia,
ana.korunovic@fon.bg.ac.rs

This paper introduces an approach for enhancing the existing web application for quantitative text analysis, QuanTA (Ljubisavljević & Devedžić, 2024), by developing *a validation module*. The main idea was to improve the quality of text analysis by validating and correcting input text through checks for profanity, grammar errors and spelling errors. As a proof of concept, the validation module was developed using the *Chain of Responsibility design pattern*.

The Chain of Responsibility design pattern avoids coupling the sender of a request to its receiver by allowing multiple objects to process the request. The objects are linked in a chain and the request is passed on until one of the objects is able to process it (Gamma et al., 1994). The general structure of the Chain of Responsibility pattern consists of three elements: *Client*, *Handler* and *ConcreteHandler* (Gamma et al., 1994; Vlajić, 2014). In this context, the structure was implemented through the definition of several classes, including *ValidationClient*, *TextValidator, ProfanityValidator*, *GrammarValidator* and *SpellingValidator*. The Handler interface (represented by the TextValidator class) defines the operation for processing a request. Each Handler contains a reference to the next handler to ensure that each object in the chain can pass the request on to the subsequent handler. ConcreteHandlers (such as ProfanityValidator, GrammarValidator and SpellingValidator) are specific implementations of the Handler interface that are responsible for processing the request.

The validation process utilizes key Python libraries, including *Better-profanity*, *Pyspellchecker* and *Gramformer*. Better-profanity is used to identify profanity words in a text, supporting the detection of leetspeak variations where letters are replaced by numbers or symbols. Pyspellchecker enables detection of spelling errors by using the Levenshtein distance algorithm and comparing words with known words from a frequency-based word list. Gramformer is developed for the detection of grammatical errors in text, using transformer-based models trained on large datasets. In addition, OpenAI's API was integrated to enable text correction.

The results related to the development are reflected in a validation module that identifies and corrects detected errors, thus preparing the text for further analysis. A graphical user interface (see Fig. 1) was developed to demonstrate the module's functionalities and ease of use. The front end was developed using React, while the back end was developed in Python and Django, with data exchange handled through the Django REST framework.

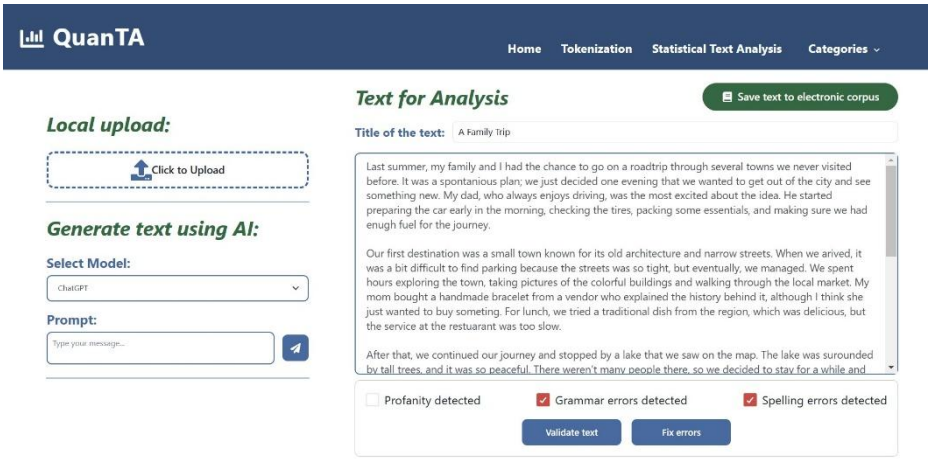


Fig. 1. Interface of the validation module

The obtained conclusions emphasize the importance of applying design patterns to create adaptable and maintainable systems for text validation. The main value of this paper is reflected in the practical application of the Chain of Responsibility pattern for text validation, providing a foundation for future system enhancements. This approach allows the system to be expanded by simply adding new validators to the chain, without the need to modify the existing code structure.

Future work will focus on extending the existing application by adding new validators and applying other design patterns. Other research directions include comparing the analysis results before and after the introduction of the text validation module in the QuanTA application to assess its impact on improving the quality of text analysis.

Keywords: Chain of Responsibility, NLP, Text Validation

References:

[1] Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns: Elements of reusable object-oriented software* (1st ed.). Addison-Wesley. https://doi.org/10.1007/978-1-4302-0096-3_10

[2] Ljubisavljević, D., & Devedžić, V. (2024). Designing a web application for text analysis. In *Universal threats in expert applications and solutions* (pp. 101–110). Springer Nature.

[3] Vlajić, S. (2014). *Софтверски патерни (Software patterns)*. Fakultet organizacionih nauka.

MSEAN: Multi-Level Semantic Enhancement with Graph Attention Network for Image-Text Matching

Jing Meng[1,2], Zhenfang Zhu[1,2], Xiangyu Shui, Guangyuan Zhang[1,2],Guang Yang[1,2], Hongli Pei[1,2]

1Shandong Key Laboratory of Technologies and Systems for Intelligent Construction Equipment, Shandong Jiaotong University, Jinan, China

2School of Information Science and Electrical Engineering, Shandong, Jiaotong University, Jinan,China mj200105@163.com, zhuzf_sdjtu@126.com,peihongli@sdjtu.edu.cn, xdzhanggy@163.com, yangguang@sdjtu.edu.cn, peihongli@sdjtu.edu.cn(Corresponding author)

Image-text matching aims to learn a common representation for images and textual descriptions, but most existing approaches either focus on global features or local region-word alignments, leading to incomplete semantic alignment [1-2]. To address this limitation, we propose Multi-level Semantic Enhancement with Graph Attention Network (MSEAN), a novel method that integrates both global and local semantics for robust cross-modal matching. MSEAN builds upon recent advances in multi-level feature learning [3-4] by introducing a joint enhancement of image and text representations using graph neural networks.

To address the limitations of isolated global or local alignment strategies in image-text matching, we propose three key innovations in MSEAN. First, we design a global–region joint learning module that integrates global image features (via ResNet) with local region-based features (from Faster R-CNN), enabling richer modeling of both region–region and region–global interactions. Second, we introduce a bilevel text feature aggregator that simultaneously captures word-level granularity and sentence-level context, thereby enhancing the completeness and coherence of textual semantics. Third, we apply Graph Attention Networks (GAT) to both visual and textual modalities to refine these multi-level features, learning salient inter- and intra-modal relationships while mitigating the influence of irrelevant components.

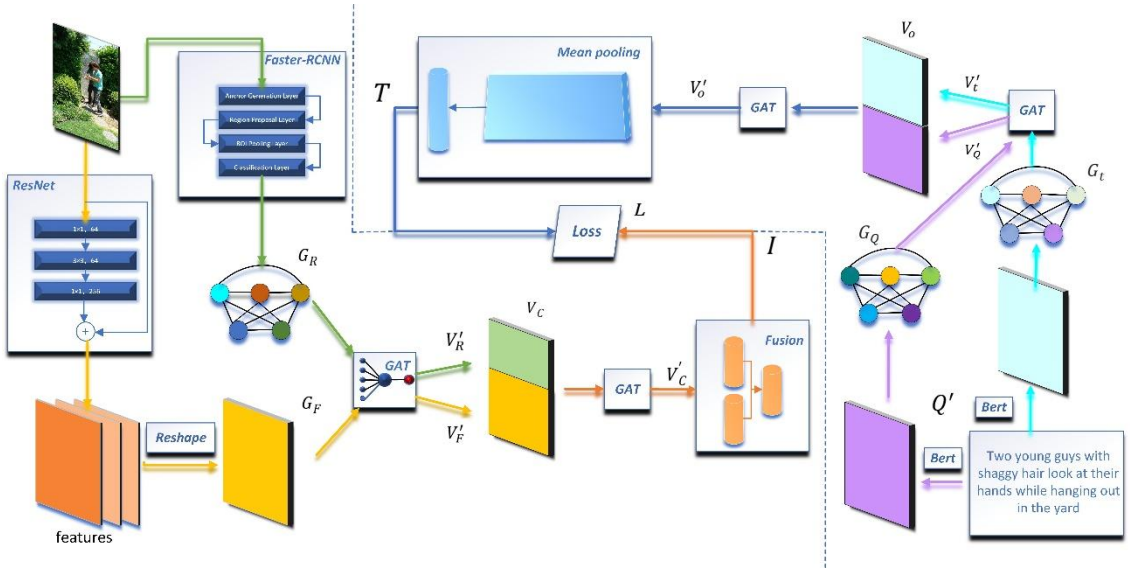


Fig. 1. Flowchart of the proposed MSEAN model. The image encoder uses a ResNet backbone for global features and a Faster-RCNN for regional features, constructing a combined global-region feature map G_R. The text encoder produces a bilevel text feature map G_Q from word embeddings and a sentence-level Bi-LSTM/BERT representation. Both image and text feature maps are further processed by Graph Attention Networks (GAT) to learn intra- and inter-modal semantic relationships. The enhanced features V_o' for image, V_t' for text are finally fused and evaluated with a contrastive loss

for matching. This multi-level architecture enables more comprehensive alignment of visual and textual content.

Results and Discussion. In experiments, MSEAN achieves new state-of-the-art results on cross-modal image–text retrieval benchmarks. On the Flickr30K dataset, our approach achieves a Recall@1 of 82.2% for retrieving the correct caption given an image, compared to 80.5% by the best previous model, and it obtains similarly strong performance on the reverse (text-to-image) retrieval task. On the larger MS-COCO dataset, MSEAN likewise attains superior results, with improved Recall@1 of 58.2% (image-to-text) and 41.8% (text-to-image) in the 5K test setting. It significantly outperforms approaches that use only global features [2] or only local alignments [3], and even surpasses the recent graph-based Dual Semantic Relations Attention Network (DSRAN) [5] in overall retrieval accuracy. These results confirm that the proposed multi-level semantic enhancement and graph attention mechanisms substantially improve the alignment of images and captions. An ablation study shows that both the global–region feature integration and the bi-level text modeling contribute significantly to the performance gains. In summary, MSEAN provides an effective framework for image-text matching, bridging the semantic gap between vision and language through multi-level feature enhancement and relational graph reasoning.

Keywords: Machine learning, Image-Text Matching, Cross-Modal Retrieval, Graph Attention Network, Visual-Semantic Alignment

References:

[1] Diao, H., Zhang, Y., Ma, L., & Lu, H. (2021). *Similarity reasoning and filtration for image-text matching*. Proceedings of the AAAI Conference on Artificial Intelligence, 35(2), 1218–1226.
[2] Faghri, F., Fleet, D. J., Kiros, J. R., & Fidler, S. (2018). *VSE++: Improving visual-semantic embeddings with hard negatives*. arXiv preprint arXiv:1707.05612.
[3] Lee, K.-H., Chen, X., Hua, G., Hu, H., & He, X. (2018). *Stacked Cross Attention for Image-Text Matching*. In Proceedings of ECCV 2018.
[4] Velicković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). *Graph Attention Networks*. International Conference on Learning Representations.
[5] Wen, K., Gu, X., & Cheng, Q. (2022). *Learning dual semantic relations with graph attention for image-text matching*. IEEE Transactions on Circuits and Systems for Video Technology.

Synergy of LLMs and ontologies to reduce cognitive load in manufacturing process

Nenad Petrovic[1], Milorad Tosic[1]

[1]Faculty of Electronic Engineering, University of Nis, Serbia.
nenad.petrovic@elfak.ni.ac.rs, milorad.tosic@elfak.ni.ac.rs

Modern manufacturing industry involves adoption of various machines and equipment, as well as human resources and skills, making the management of the underlying processes increasingly difficult. Therefore, planning, supportive and decision-making activities are becoming more and more complex, considering both the scale and heterogeneity of the elements involved. State-of-art information systems aim to provide means and capabilities which target the automation of those activities. One of possible solutions is utilization of Semantic Knowledge Graphs (SKG) in order to represent knowledge about crucial aspects of the underlying manufacturing-related processes (Tosic, 2024). It provides supportive mechanisms for retrieval of relevant data together with a context required for decision-making using queries and reasoning against semantic representation of information, so called intelligent data, within the SKG (Tosic, 2025).

In this paper, we examine potential of adopting synergy of SKGs and Large Language Models (LLMs) in order to increase degree of automation, while in the same time reduce the cognitive load in semantics-based manufacturing process management solutions. Often, similar works do not address the issue of handling larger text inputs, making them less effective in practice (Cremaschi, 2025). While some of them adopt Retrieval Augmented Generation (RAG) for handling large user-provided freeform documents as inputs (Gauthier, 2025), most of them do not consider the cases when semantic data used to provide context also requires large amount of text. For that reason, we aim to bridge this gap by making use of Retrieve and Re-Rank technique in order to extract relevant parts from ontologies and knowledge graphs, with a focus on two use cases within the context of manufacturing: 1) question-answering agent about the content of ontologies 2) semantic annotation of freeform text with respect to given ontology. Overview of the proposed workflow is given in Fig. 1.

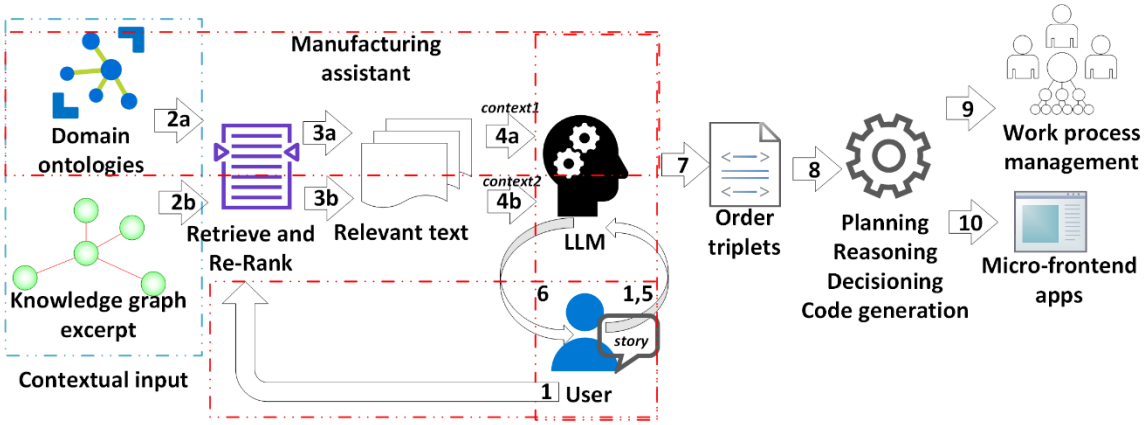


Fig. 5 LLM-aided semantic-enabled manufacturing workflow: 1,5-User input{story} 2a-RDF schema 2b-RDF knowledge graph 3a-Ontolgy chunks 3b-Knowledge graph excerpt chunks 4a-Context: ontology excerpt {context1} 4b-Context: knowledge graph template {context2} 6-Intermediary result 7-Triplets 8-Order definition 9-Generated work plans 10-Generated micro-frontends

The process begins with the user providing a freeform textual description of relevant aspect from manufacturing domain, such as orders, underlying activity flow, involved machines, employees, and other operational details (Tosic, 2024). This input, referred to as the *story*, serves as the starting point for the Retrieve and Re-Rank process (Gupta, 2018). As illustrated in Fig. 1, this process identifies and extracts the most relevant segments from the ontology collection, producing context1 (denoted as

4a) which is an ontology excerpt tailored to the user's input. To enhance precision and supress hallucinations by the LLM, an additional excerpt (denoted as 4b – context2) is retrieved from a representative knowledge graph. This dual-context approach ensures that the terminology and structure align closely with the intended domain semantics. While the Retrieve and Re-Rank operations are performed locally on the host machine (multi-qa-MiniLM-L6-cos-v1 as bi-encoder and cross-encoder/ms-marco-MiniLM-L-6-v2), while LLM used for response generation (GPT-4o) relies on OpenAI's cloud. The initial prompt, constructed in step 5, combines the user *story* and the retrieved context. Here, *context1* and *context2* represent the outputs of the Retrieve and Re-Rank process, while *story* is the user-provided input. The generated RDF triplets describe the order in a form suitable for insertion into a semantic knowledge graph. The workflow also supports interactive „human-in-the-loop" style - users can review intermediate results and refine them through additional feedback (based on prompt 2, so flow repeat steps 5 and 6). The resulting knowledge graph further serves as input to semantics-based manufacturing support tools, providing capabilities of reasoning, planning, decision-making, as well as further steps of code generation based on semantic representation. The outcomes of such steps can either affect the management of manufacturing process or lead to creation of supportive applications based on semantic micro-frontends (Tosic, 2024). In parallel, the portion of the system highlighted by the red boundary in Fig. 1 enables the LLM-based planning assistant. Overview of prompts and experiments based on 10 runs is given in Table 1. Based on outcomes, the approach seems promising when it comes to acceleration and enhanchement of semantics-driven processes.

Table 1 Prompts and experiments overview

Prompt	Scenarios	Example	Result
Prompt1: Create set of RDF triplets for semantic knowledge graph in XML with respect to given ontologies: {context1 – ontology excerpt} and example graph {context2 – knowledge graph template} based on user story: {story}"	Order creation, employee creation, machine creation, activity flow definition.	We have a new order for C1 company and the product we want to produce is P1. The activity starts from 2025-07-01 and ends 2025-08-31.	1/1 class and 5/5 relations in 80% cases
Prompt2: Update set of RDF triplets for {current result}".	Employee update	Change id of employee Nenad Petrovic.	Correct update in 100% cases
	Flow extension	Add new activity: preparation – before cutting activity to given flow.	4/4 classes, 3/3 relations, no hallucinations in 70% cases
Prompt3: Answer the question about ontology: {story}, based on given excerpt: {context1 – ontology excerpt}"	Related concepts retrieval	Which concepts are related to order?	3/3 in 80% cases
	Attribute list retrieval	Which elements are relevant for composite activity	4/4 in 90% cases
	Relations retrieval	How empoloyee is related to order	Correct in 70% cases

Keywords: Ontology, Semantics, Large Language Model (LLM), Retrieval Augmented Generation (RAG), Planning

References:

[1] Tosic, M., Petrovic, N., & Tosic, O. (2024). Wearable networks for semantics-driven FASPAS approach to fatigue management. In Proceedings of BalkanCom 2024 (pp. 75–80). IEEE.
[2] Tosic, M., Petrovic, N. (2025). Intelligent data platform – one step at a time, Artificial Intelligence conference, Serbian Academy of Sciences and Art.
[3] Cremaschi, M., D'Adda, F., & Maurino, A. (2025). StEELlm: An LLM for generating semantic annotations of tabular data. ACM Transactions on Intelligent Systems and Technology.
[4] Gauthier, J.-M., Jenn, E., & Conejo, R. (2025). Ontology-driven LLM assistance for task-oriented systems engineering. In Proceedings of MODELSWARD 2025 (pp. 383–394). SCITEPRESS.
[5] Gupta, V., Chinnakotla, M., & Shrivastava, M. (2018). Retrieve and re-rank: A simple and effective IR approach to simple question answering over knowledge graphs. In Proceedings of the 1st Workshop on Fact Extraction and VERification (FEVER) (pp. 22–27). Association for Computational Linguistics.

On the Robustness and Privacy of Lightweight LLM Routers in Retail Contexts

Sandi Baressi Šegota[1], Nikola Anđelić[2], Vedran Mrzljak[3], Ivan Lorencin[4]

[1]University of Rijeka, Faculty of Engineering, sandi.segota@riteh.uniri.hr
[2]University of Rijeka, Faculty of Engineering, nikola.andelic@riteh.uniri.hr
[3]University of Rijeka, Faculty of Engineering, vedran.mrzljak@riteh.uniri.hr
[4]Juraj Dobrila University of Pula, Faculty of Informatics, ilorencin@unipu.hr

This study explores the application of lightweight large language models (LLMs) as intelligent routers in retail customer support scenarios. We have performed an evaluation of multiple lightweight and commercial LLMs, such as LLaMA3, DeepSeek, and GPT, across various customer intent categories, including task-specific classification and response triage. Routing decisions in such systems are critical to ensuring users are connected to the correct service or information channel with minimal delay.

The experimental setup involved seven retail-specific intent classes: *product_search*, *order_status*, *returns_and_refunds*, *store_locator*, *promotions_and_deals*, *connect_to_agent*, and *general_info*. Each model was evaluated based on the number of correctly classified prompts per category. The results, presented in Figure 1, indicate varying performance between the models depending on the task type.

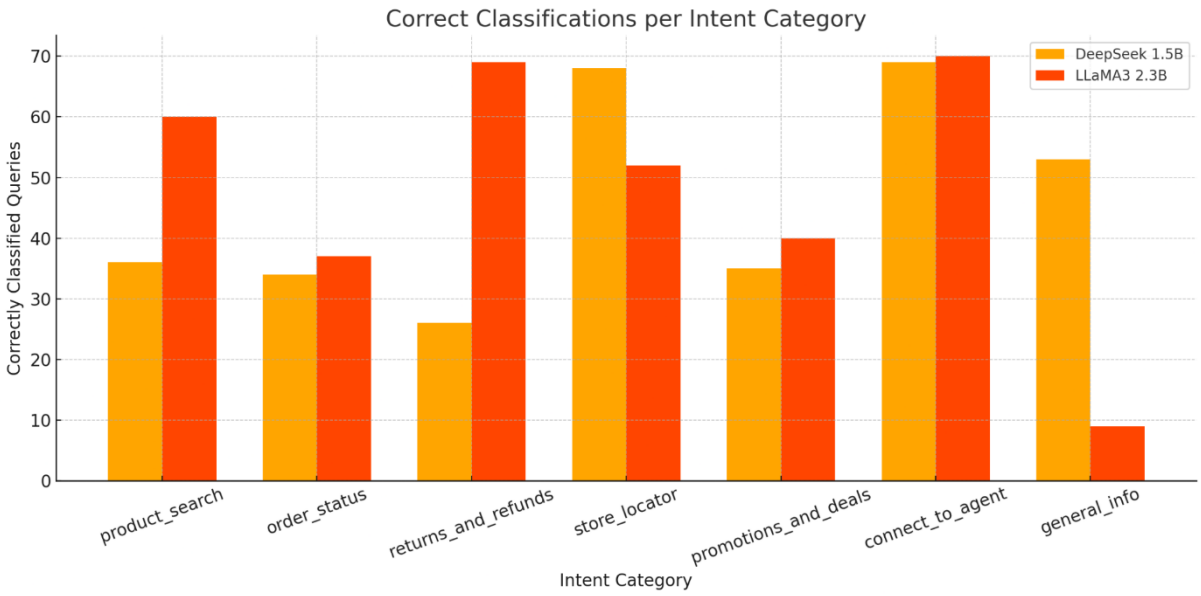


Figure 1. Comparison of correctly classified customer queries across intent categories by DeepSeek 1.5B and LLaMA3 2.3B. The results show varying model performance depending on query complexity, with LLaMA3 outperforming DeepSeek in complex intents such as *product_search* and *returns_and_refunds*, while both models perform similarly in straightforward categories like *store_locator* and *connect_to_agent*.

When the results are compared across intent categories, DeepSeek 1.5 B and LLaMA3 2.3 B display comparable performance in “store_locator” and “connect_to_agent”, but LLaMA3 notably outperforms DeepSeek in “product_search” (60 vs. 36) and “returns_and_refunds” (69 vs. 26). This suggests that LLaMA3—when deployed locally—achieves high accuracy while offering a strong privacy benefit, as outlined by Kassem, Schölkopf & Jin (2025) in *How Robust Are Router-LLMs?* Ensuring customer data remains within the retail infrastructure mitigates compliance risks under GDPR and CCPA.

These outcomes align with Alexander & de Vries (2025), who compare weak supervision and LLMs for short-query classification, and with Zhang et al. (2024), who present embedding-based fine-tuning approaches for intent classification and out-of-scope detection. Combining local LLaMA deployment with embedding-based reranking presents a practical and scalable solution that balances performance with regulatory compliance. Future work will explore ensembling methods and hybrid architectures to combine strengths of multiple lightweight LLMs (Zhang et al., 2024).

Keywords: Large Language Models, Retail AI, Intent Classification, LLM Routing, Customer Support

References:

[1] Kassem, A. M., Schölkopf, B., & Jin, Z. (2025). How Robust Are Router-LLMs? Analysis of the Fragility of LLM Routing Capabilities. arXiv preprint arXiv:2405.04734. https://arxiv.org/abs/2405.04734

[2] Alexander, D., & de Vries, A. P. (2025). In a Few Words: Comparing Weak Supervision and LLMs for Short Query Intent Classification. arXiv preprint arXiv:2503.08504. https://arxiv.org/abs/2503.08504

[3] Zhang, T., Norouzian, A., Mohan, A., & Ducatelle, F. (2024). A new approach for fine-tuning sentence transformers for intent classification and out-of-scope detection tasks. arXiv preprint arXiv:2410.13649. https://arxiv.org/abs/2410.13649

Pilot Text to Text Transfer Transformer Model for Serbian Language

Mihailo Škorić[1]

[1]University of Belgrade, Serbia, mihailo.skoric@rgf.bg.ac.rs

In this presentation, we will introduce the first encoder-decoder language model trained specifically for Serbian and built upon T5 (Text-to-Text Transfer Transformer) architecture (Raffel et al, 2020). This architecture is developed by Google in pursuit of standardization of sequence-to-sequence language modelling, as the continuation of their work on the initial transformers architecture (Vaswani et al, 2017). The aim of this research is to provide the initial T5 model for Serbian, which will be the first in the series that will handle multiple natural language processing tasks such as text summarization, question-answering or even machine translation.

Since the idea is to build a model designed specifically for Serbian, the model is trained using large textual datasets in Serbian only consisting of over nine billion words (Škorić & Janković, 2024): Serbian subset of *Kišobran*, the umbrella web corpus of the Serbo-Croatian macro language (8.1 billion words), Serbian subset of corpus *Znanje*, containing over eleven thousand doctoral dissertations and over eleven thousand research papers written in Serbian (700 million words), and finally Serbian subsets of *Vikipedija* and *Vikizvornik* corpora produced from Serbian Wikipedia and Wiki source dumps (150 and 20 million words respectively).

The specific architecture of the model is based on *flan-t5-base* (248 million parameters), which consists of both an encoder and a decoder, following a text-to-text approach for all natural language processing tasks. The model uses multi-head self-attention and feed-forward layers optimized for sequence generation, so, unlike BERT, which does not have a decoder component, T5 has the ability to generate natural text outputs e.g. an answer to a question or a translation of the input text. The architecture also ensures flexibility in fine-tuning for a multitude of natural language processing tasks: text correction, summarization, style transfer and question-answering which are very relevant for the Serbian language as the solutions for these tasks are quite sparse. Additionally, the models trained on these tasks could be further fine-tuned to tackle some more Serbian-specific tasks, such as Latin script diacritic restauration and interdialectal text transformation e.g. from Ekavian to Iekavian.

The model is trained using text correction as the primary objective. To train the model for said task, we employ a dataset of grammatically correct Serbian sentences as output (previously mentioned corpora), while generating corrupted versions as input through programmatic transformations. These distortions include diacritics and punctuation removal, grammatical inconsistencies, missing words, and misplaced characters, reflecting common mistakes found in real-world data. The approach is inspired by Noisy Channel approaches in NLP, which model corrections by learning from distorted input-output pairs (Brill & Moore, 2000). By training in this manner, the model learns to generate correct Serbian sentences when given noisy text, or to output the inputted text if there is no need for correction.

Evaluation is conducted on text summarization task, using a specially prepared dataset for fine-tuning and evaluation. We assess performance using Recall-Oriented Understudy for Gisting Evaluation metric (ROUGE) score, comparing a total of eight candidate models against multilingual baseline of similar size (300 million parameters), *mT5-small* (Xue et al, 2020). The results show that our best monolingual Serbian T5 model outperforms the multilingual model despite smaller size, achieving moderate ROUGE scores (~0.30). Future work will focus on refining text corruption strategies for the purpose of more rigorous training, training fine-tuned version for more specific NLP tasks and exploring larger model variants, such as *flan-t5-large* to further evaluate the architecture's performance for Serbian text processing.

Keywords: Language models, Serbian, T5, Text correction, summarization

References:

[1] Brill, E., & Moore, R. C. (2000, October). An improved error model for noisy channel spelling correction. In *Proceedings of the 38th annual meeting of the association for computational linguistics* (pp. 286-293).

[2] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, *21*(140), 1-67.

[3] Škorić, M., & Janković, N. (2024). New Textual Corpora for Serbian Language Modeling. *Infotheca: Journal for Digital Humanities*, *24*(2).

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

[5] Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., ... & Raffel, C. (2020). mT5: A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv:2010.11934*.

Developing Language Resources for Recognizing Moral Aspects in the Serbian Language: Multi-label Categorization through the Perspective of Moral Foundations Theory

mr Milena Šošić[1], prof. dr Ranka Stanković[2], prof. dr Jelena Graovac[3]

[1]PhD candidate, Department for Computer Science, Faculty of Mathematics, University of Belgrade, Serbia, pd202030@alas.matf.bg.ac.rs, milena.sosic@gmail.com
[2]Departman for Applied Mathematics and Informatics, Faculty of Mining and Geology, University of Belgrade, Serbia, ranka.stankovic@rgf.bg.ac.rs
[3]Department for Computer Science, Faculty of Mathematics, University of Belgrade, Serbia, jelena.graovac@matf.bg.ac.rs

Through this study, we present the development process of an annotated corpus, a lexicon of moral words, and advanced machine learning models for recognizing the moral aspect of Serbian language using a multi-label annotation approach based on Moral Foundations Theory – MFT (Graham, 2009). Due to the lack of available linguistic resources for specific language aspects for languages like Serbian that exist for other languages (Hoover, 2020), this study aims to expand the capabilities for analyzing moral discourse through the application of advanced computational linguistics technologies. Constructed *Social-Moral.SR* corpus contains ~13.6k conversational messages collected from social media platforms X and Reddit written in Serbian language and annotated for moral categories. The development of the corpus began with the automatic collection and keyword-based selection of the messages, followed by pre-annotation using advanced language models that understand the Serbian language. The selected advanced model, *Falcon-7b-Instruct*, with such characteristics, enabled the initial classification of messages obtained from social networks. The categorization was performed according to the ten categories of moral sentiment defined by MFT[6], such as: care/harm, fairness/cheating, loyalty/betrayal, authority/subversion and purity/degradation. A critical challenge in pre-annotation lay in adapting the model to the Serbian language and culture, which required additional manual verification and adjustments to ensure the relevance of the obtained moral categorizations (He, 2024; Zangari, 2025). The evaluation of the verified annotations was performed using Cohen's Kappa coefficient as the main measure to verify the degree of agreement between the annotators. The reliability coefficient, with an average value of 0.36, indicates a moderate but acceptable level of agreement among annotators for the given task and the number of categories in the established annotation schema.

Based on the annotated *Social-Moral.SR* corpus, the lexicon of moral words in Serbian language, named *MFD.SR* encompassing in total ~4.3k lemma-PoS pairs, was developed automatically using a combination of natural language processing (NLP) resources already developed for Serbian language such as PoS taggers, lemmatizers, NER taggers and sentiment intensity evaluators (SRPOL). To systematically identify characteristic moral words for each category, a class-based Tf-Idf algorithm was applied. This approach facilitates the extraction of words that are characteristic of each moral foundation. The verification of the lexicon developed using this approach was conducted through a survey that assessed the understanding of moral foundations among anonymous participants from the Serbian-speaking community. Metrics of statistical significance (F-statistics) and correlation (Pearson coefficient), measured on the number of recognized moral words on textual descriptions obtained from the survey answers, indicate the high level of statistical significance and correlation on each moral category, which confirms the lexicon's ability to appropriately recognize moral words in the written texts.

Moreover, advanced machine learning models *Moral-BERT.SR* and *Moral-LLaMA.SR* were developed as a result of fine-tuning different distributions of pre-trained BERT (multiple model versions) and

[6] https://moralfoundations.org/

LlaMA (model *LlaMA-3.2-3B-Instruct*) language models using the annotated *Social-Moral.SR* corpus as a training source. With this approach, fine-tuned base models are designed to recognize moral categories in texts from social networks, with multi-label annotation objective set in the task of recognizing moral signals in textual content. *Moral-BERT.SR* and *Moral-LLaMA.SR* achieved F_1 score of ~68% and ~56% respectively, which are in the range with the results published in other studies on the same task (Bulla, 2022). The *Moral-LLaMA.SR* showed significant improvement in the performance compared with the *LlaMA-3.2-3B-Instruct* base model on the moral classification task of Serbian messages (zero-shot technique), indicating the necessity for fine-tuning the model for the task and language in use. The limitations observed with the fine-tuned LLaMA model underscore the current gaps in Serbian language support and the necessity to train the models on larger corpora with a greater number of parameters for enhanced performance. Developed models significantly improve the accuracy and understanding of morality within the Serbian language compared to the lexicon-based model, which make them valuable tools for further research and analysis of moral content in conversational texts in the Serbian discourse.

This research represents a significant step towards developing new linguistic resources for the Serbian language in the context of moral psychology and computational linguistics. Despite numerous challenges, the developed linguistic resources provide a solid foundation for future research and applications, including moral reasoning analysis, moral sentiment recognition, and support for sociological studies in the Serbian language. Future work will focus on enriching and verifying the corpus, enhancing the models of automatic moral stance recognition, and refining methods for extracting moral words and phrases from texts. These efforts aim to increase the effectiveness of the developed linguistic resources for analyzing the moral aspects of the Serbian language.

Key contributions: (1) first Serbian corpus annotated for moral categories grounded in Moral Foundations Theory, (2) first automatically constructed moral lexicon for Serbian derived from social media discourse, and (3) novel fine-tuned language models for multi-label moral text classification in Serbian.

Keywords: moral, conversational texts, annotation, classification, linguistic resources, Serbian language

References:

[1] Bulla, L. a. (2022). Detection of morality in tweets based on the moral foundation theory. *International Conference on Machine Learning, Optimization and Data Science*, 1--13.

[2] Graham, J. a. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology, 96*, 1029.

[3] He, Z. a. (2024). Whose Emotions and Moral Sentiments Do Language Models Reflect? *Findings of the Association for Computational Linguistics ACL 2024*, 6611--6631.

[4] Hoover, J. a.-W. (2020). Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science, 11*(8), 1057--1071.

[5] Zangari, L. a. (2025). A survey on moral foundation theory and pre-trained language models: current advances and challenges. *AI & SOCIETY*, 1--26.

From Unstructured Interviews to Queryable Knowledge: An AI Pipeline for the "Digitalne Ikone" Corpus using TEI, NER, NEL, and RAG

Ranka Stanković[1], Tamara Vučenović[2], Milica Ikonić Nešić[3]

[1]University of Belgrade, Faculty of Mining and Geology, Serbia, ranka@rgf.rs
[2]University Metropolitan, Faculty of Management, tamara.vucenovic@metropolitan.ac.rs
[3]University of Belgrade, Faculty of Philology, Serbia, milica.ikonic.nesic@fil.bg.ac..rs

Layers of structure and semantic annotation are added to the raw text through a series of sequential stages in the transformation process. Transcripts in plain text were first converted to the Text Encoding Initiative (XML TEI) format. This included identifying and marking up fundamental elements, such as interview boundaries, paragraph breaks (<p>), and speaker turns (<speaker/interviewee>) with the identifier for each speaker; dates of interview with <month>, <year> tags; topic with <title>,... This structural encoding is critical, as it segments the text logically, providing essential context boundaries for subsequent Natural Language Processing (NLP) tasks and enabling querying based on structural elements (e.g., retrieving content spoken by a specific person). The TXM tool (https://txm.gitpages.huma-num.fr/textometrie/) was used for a basic NLP pipeline and textometry with Jerteh resources (https://jerteh.rs) (Stanković et al. 2020), while NER/NEL models were provided from the Text Embeddings - Serbian Language Applications (TESLA) project (Ikonić Nešić et al 2024).

Following structural encoding, an automated Named Entity Recognition (NER) process was applied to the TEI documents. Utilizing NLP models, occurrences of key Named Entities (NEs) within the text were identified and classified. Entity types of particular relevance for this corpus include persons (<PERS>), places (<LOC>), organizations (<ORG>), and events (<EVENT>). The results of the NER process were integrated into the TEI documents by wrapping the identified text spans with the corresponding TEI tags. This step created an inventory of potentially significant entities mentioned throughout the interviews, automating a fundamental step in semantic analysis.

The critical task of linking the identified entity mentions (from NER) to canonical, unambiguous entries in external Knowledge Bases (KBs) was tackled in the third stage, Named Entity Linking (NEL). Because of its extensive coverage and organized format, Wikidata was selected as the main target KB. To resolve ambiguity and link each entity mention to its unique identifier in Wikidata, the NEL process involved analyzing the context surrounding each mention. The NEL process involved analyzing the context surrounding each entity mention to resolve ambiguity and linking the mention to its unique identifier in Wikidata. The results of the NER process were integrated into the TEI documents by wrapping the identified text spans with the corresponding TEI tags. This step created an inventory of potentially significant entities mentioned throughout the interviews, automating a fundamental step in semantic analysis (Ikonić Nešić et al. 2022). This NEL step significantly enriches the corpus semantically. This not only aids disambiguation but also implicitly links the interview content to the vast network of facts and relationships available in Wikidata, enabling future inference and data integration. This stage effectively transforms the text-based entity mentions into semantic links within a larger knowledge graph paradigm. Some of the interviewers were present in Wikidata (e.g. Voja Antonić https://www.wikidata.org/wiki/Q1252236), but several were missing, so we had to enrich Wikidata with the required information.

The culmination of the pipeline is the implementation of advanced QA capabilities based on the structured and semantically enriched TEI corpus. Instead of relying solely on the generative capabilities of LLMs (which can suffer from hallucination and lack of grounding in specific sources), we adopt a Retrieval Augmented Generation (RAG) framework. This approach combines the strengths of information retrieval and language generation, where retrieved passages are provided as context to the LLM with a task to synthesize the information from the provided context. This RAG approach ensures that the generated answers are grounded in the actual content of the "Digitalne

Ikone 20+" (Vučenović 2024) interviews, reducing the risk of factual errors or irrelevant outputs. LLMs, specifically instructed through the prompt engineering framework DSPy (Declarative Self-improving Python, https://dspy.ai/), are employed in the generation phase to produce fluent and informative responses based on the retrieved evidence.

This pipeline demonstrates a powerful methodology for transforming unstructured textual archives into valuable, queryable knowledge resources accessible through AI-driven interfaces.

This means making it possible for researchers, students, and the general public to engage with the interviews in previously unattainable ways for the "Digitalne Ikone" corpus, allowing for intelligent summarization, automated fact extraction, and semantic search. In particular, the underlying knowledge representation may be further enhanced by graph-based methods that directly predict missing entities, classes, or links within a Knowledge Graph generated from the corpus and its NEL links. Further work involves investigating more complex RAG architectures and improving NER/NEL's accuracy for the particular domain language of the interviews.

The "Digitalne Ikone" interview corpus has been successfully transformed into a semantically enriched and structured representation by putting in place a pipeline that proceeds from plain text through TEI encoding, NER, and NEL. This provides a strong basis for creating sophisticated QA systems. The corpus is transformed into a dynamically queryable knowledge base through the use of the RAG framework, which is powered by LLMs and allows for grounded, coherent answers to complex queries. This work greatly improves the accessibility and knowledge-discovery potential of similar text collections by offering a replicable model for applying contemporary AI and NLP techniques to them.

Keywords: Named Entity Linking, Information Extraction, Textometry, Knowledge Graphs, RAG

References:

[1] Ikonić Nešić, M., Stanković, R., Schöch, C., & Škorić, M. (2022). From ELTeC text collection metadata and named entities to linked-data (and back). U Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference, str. 7-16. European Language Resources Association (ELRA).

[2] Ikonić Nešić, M., Petalinkar, S., Stanković, R., Utvić, M., & Kitanović, O. (2024). SrpCNNeL: Serbian model for named entity linking. U 2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS), str. 465-473. IEEE.

[3] Nikolić, M., & Vučenović, T. (2024). Veštačka inteligencija (VI) u kulturi i medijima - od distopije do utopije i nazad. Kultura, 184, 5-19. https://doi.org/10.5937/kultura2484005N

[4] Stanković, R., Šandrih, B., Krstev, C., Utvić, M., & Škorić, M. (2020). Machine learning and deep neural network-based lemmatization and morphosyntactic tagging for Serbian. U LREC 2020-12th International Conference on Language Resources and Evaluation, Conference Proceedings, str. 3954-3962. European Language Resources Association (ELRA).

[5] Vučenović, T. (2024). Digitalne ikone 20+. Akademska knjiga.Vučenović, T. (2024). Digitalne ikone 20+. Akademska knjiga.

Large-Scale Simulation of Reddit Communities Using Multi-Agent LLM Systems

Aleksandar Tomašević[1], Darja Cvetković[2], Sara Major[1], Slobodan Maletić[3], Miroslav Anđelković[3], Ana Vranić[2], Boris Stupovski[2], Dušan Vudragović[2], Aleksandar Bogojević[2], Marija Mitrović Dankulov[2]

[1]Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia, atomashevic@ff.uns.ac.rs
[2]Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia
[3]Vinča Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia

Large Language Models (LLMs) present unprecedented opportunities for simulating complex online social dynamics at scale. This paper presents a comprehensive multi-agent simulation of Reddit-style technology discussion communities using an adapted version of the Y Social framework (Rossetti et al. 2024), originally designed for Twitter simulations but forked and modified for Reddit-like environments.[7] We specifically target the simulation of r/technology communities because technology topics often contain contentious subjects that are more or less politically polarizing - from Big Tech regulation to AI ethics to Elon Musk's ventures - allowing us to study how politically charged and potentially toxic conversations emerge within communities not explicitly focused on politics. Our simulation encompasses one month of activity with approximately 50,000 posts and comments generated by 7,637 unique LLM agents.

The simulation integrates real-world content through RSS feeds from 33 diverse technology news sources from TechCrunch and Wired to Fox News Technology and CNN Tech. This approach ensures that agent discussions are grounded in authentic current events rather than artificial scenarios. Simulation parameters were calibrated using data drawn from the MADOC dataset (Dankulov et al. 2025), ensuring realistic behavioral patterns and community dynamics. Each agent is powered by the uncensored Dolphin3 LLM (based on Llama 3.1 8B) and possesses detailed psychological profiles including demographic characteristics, political affiliations, Big Five personality traits, education levels, and critically, configurable toxicity propensities ranging from "absolutely no" to "extremely" toxic behavior.

Our methodology employs prompt engineering to enable agents to engage in contextually appropriate behaviors including posting original content, commenting on news articles, participating in threaded discussions, and exhibiting realistic activity patterns across 24-hour cycles. The system successfully generated conversations demonstrating sustained multi-turn interactions characteristic of real online communities. Figure 1 demonstrates key emergent properties: (left) power-law distribution in posting behavior with 7,637 agents showing patterns that mirror empirical observations from actual social media platforms, and (right) k-core network structure revealing the community's hierarchical organization, with high k-core nodes (yellow/green) forming the dense interaction core and lower k-core nodes (blue/purple) positioned at the periphery.

[7] https://github.com/atomashevic/YClient-Reddit

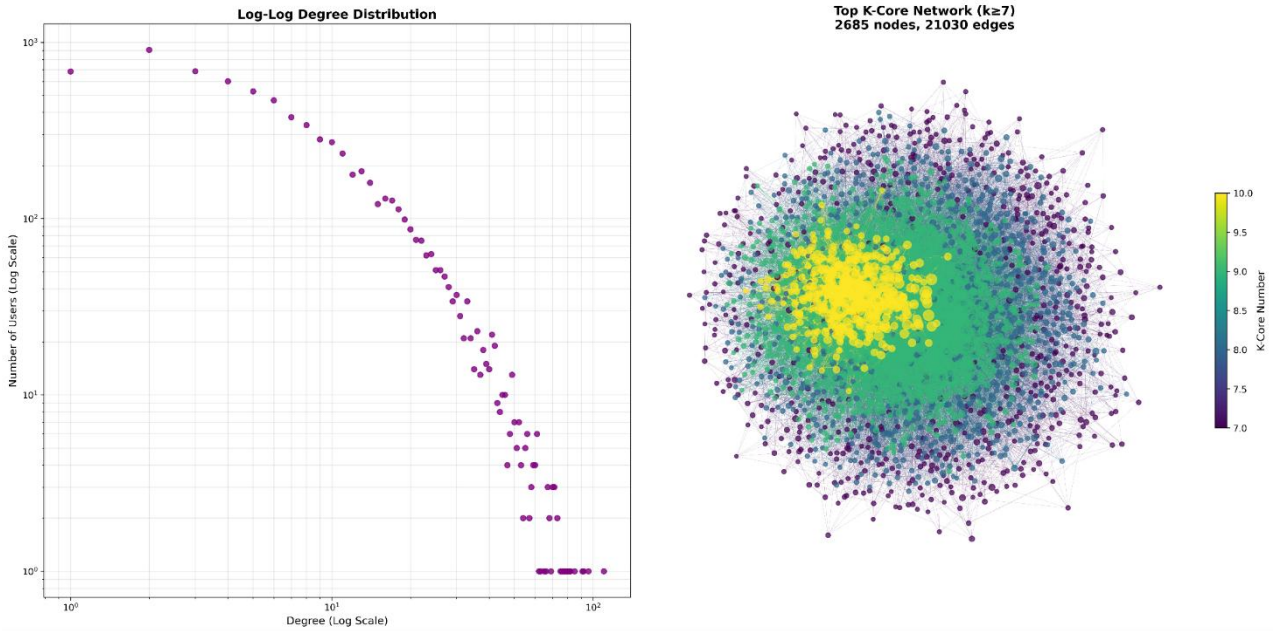


Fig. 1 Left panel: Log-log degree distribution showing power-law posting degree patterns across 7,637 agents. Right panel: Top k-core network visualization displaying emergent hierarchical community structure, where node colors represent k-core values and edges show interaction patterns between high-influence agents.

Toxicity analysis reveals compelling emergent patterns: 10.55% of content exceeded moderate toxicity thresholds (0.5), with 17.97% showing elevated toxicity levels (0.25+), closely matching empirical observations from actual technology forums. We demonstrate fidelity between our simulation and real sample of Reddit r/technology data, with both distributions exhibiting similar right-skewed patterns dominated by low-toxicity content (>80% below 0.2), comparable tail behaviors for moderate-to-high toxicity ranges and similar mean toxicity scores (0.1327 in simulated community vs. 0.1339 in real r/technology sample).

This work advances AI research in several domains: multi-agent systems capable of sustained social simulation, prompt engineering for realistic behavioral modeling, and computational social science methodologies. The framework provides researchers with a controlled environment for studying online community dynamics, testing moderation strategies, and understanding the emergence of toxic discourse patterns without ethical constraints of live platform experimentation. Our results demonstrate that LLM-based agent societies can serve as valuable digital laboratories for investigating complex social phenomena at unprecedented scale and fidelity.

Keywords: Large Language Models, Multi-Agent Systems, Social Simulation, Online Communities

References:

[1] Dankulov, M. M., Tomašević, A., Maletić, S., Anđelković, M., Vranić, A., Cvetković, D., Stupovski, B., Vudragović, D., Major, S., & Bogojević, A. (2025). Multi-Platform Aggregated Dataset of Online Communities (MADOC). In arXiv [cs.SI]. arXiv. http://arxiv.org/abs/2501.12886

[2] Rossetti, G., Stella, M., Cazabet, R., Abramski, K., Cau, E., Citraro, S., Failla, A., Improta, R., Morini, V., & Pansanella, V. (2024). Y Social: an LLM-powered Social Media Digital Twin. In arXiv [cs.AI]. arXiv. http://arxiv.org/abs/2408.00818

Active Learning in Automatic Speech Recognition

Vladimir Vincan[1], Ognjen Kundačina[1], Dragiša Mišković[1]

[1] The Institute for Artificial Intelligence Research and Development of Serbia,
vladimir.vincan@ivi.ac.rs, ognjen.kundacina@ivi.ac.rs, dragisa.miskovic@ivi.ac.rs

Building high-quality automatic speech recognition (ASR) systems requires large amounts of accurately labeled data, yet manual transcription remains a time-consuming and resource-intensive process, especially for low-resource languages such as Serbian. While existing high-quality corpora for Serbian are primarily based on read speech, which is typically collected in controlled environments, our efforts have focused on transcribing spontaneous spoken interactions that better capture the natural variability found in real-world ASR applications. This shift addresses the domain mismatch between traditional datasets and the distributional characteristics of conversational speech, which is often more informal, disfluent, and acoustically diverse. In our ongoing work, we have focused on expanding the quantity and diversity of labeled speech data in Serbian, resulting in around 50 hours of newly annotated audio.

Active Learning (AL) encompasses a range of machine learning strategies designed to identify and prioritize the most informative samples from a large pool of unlabeled data. In the context of ASR, vast amounts of untranscribed audio are readily available, yet manually labeling all of it is impractical due to time and resource constraints. AL presents a compelling solution by enabling the selection of only the most valuable segments for annotation, thereby maximizing model improvement while minimizing labeling effort. In this work, we propose an unsupervised AL approach that operates without requiring model inference over an initial labeled dataset – a step commonly found in supervised AL frameworks. This design reduces computational overhead and simplifies the pipeline, making the method well-suited for low-resource settings and data expansion. Our aim is to use this algorithm for future Serbian ASR dataset expansion.

Unsupervised AL has been previously explored in ASR, with one of the most widely used methods being k-means clustering applied to i-vector embeddings (Malhotra, 2019). While this approach has demonstrated some success, it comes with notable limitations. First, it introduces additional hyperparameters to balance uncertainty and diversity during sample selection, increasing the complexity of tuning the method effectively. Second, the i-vectors algorithm (Dehak, 2011), which has been originally developed for speaker classification, is no longer considered state-of-the-art, limiting their effectiveness in capturing the variability relevant for ASR. Furthermore, the k-means algorithm partitions the embedding space into a fixed number of compact clusters, assigning each sample to its nearest centroid. This hard clustering strategy may obscure rare or underrepresented speech patterns, thereby reducing the method's ability to identify informative and diverse samples that would benefit model training the most.

To address the limitations of previous approaches, we propose an unsupervised AL method that leverages the embedding layer of a state-of-the-art x-vector speaker classification model (Snyder, 2018), followed by clustering using the DBSCAN algorithm (Schubert, 2017). X-vector embeddings have been shown to achieve better results than i-vectors on datasets with diverse speakers, indicating improved discriminability and robustness for speaker-related tasks. Unlike k-means, DBSCAN does not require the number of clusters to be predefined and is well-suited for discovering clusters that include smaller, potentially underrepresented groups. To further enhance sample diversity and representation, we apply disproportionate cluster sampling, prioritizing a slightly higher proportion of samples from smaller clusters. For a desired size of the selected subset $|S|$ and for a known number of clusters K and size of each cluster $|C_k|$, we can define a cluster-dependent coefficient that assigns higher values to smaller clusters, called the scaling factor $\alpha_k(|C_k|)$. The scaling

factor is implemented as a linearly decreasing function, mapping the sampling rate of cluster C_k in S to the relative size of the cluster $|C_k|$. The formula can be written as follows:

$$|S_k| = \left\lceil \alpha_k(|C_k|) \cdot \frac{|C_k|}{\sum_{i=1}^{K} |C_i|} \cdot |S| \right\rceil ,$$

(1)

This formula allows for more balanced data selection, increasing the likelihood of capturing atypical or rare speech and acoustic patterns that are often missed in traditional clustering approaches. To the best of our knowledge, the proposed approach is the first algorithm that performs AL using x-vectors clustering, with an additional contribution in disproportionate cluster sampling.

We evaluated our proposed AL strategy by fine-tuning a 300-million parameter Wav2Vec2 XLS-R model (Babu, 2022) on a labeled subset selected from a larger pool of unlabeled data. The model was trained under consistent conditions and hyperparameters, and performance was assessed on both in-distribution and out-of-distribution test sets. Our results, measured using the word error rate metric, demonstrate that the proposed approach outperforms both random sampling and the traditional i-vector with k-means clustering method. These findings highlight the effectiveness of leveraging x-vector embeddings with density-based clustering for selecting informative and diverse samples in low-resource ASR settings.

Keywords: Automatic speech recognition, active learning

References:

[1] K. Malhotra, S. Bansal, and S. Ganapathy (2019). *Active Learning Methods for Low Resource End-to-End Speech Recognition*. In Proc. Interspeech 2019 pp. 2215–2219

[2] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet (2011). Front-End Factor *Analysis for Speaker Verification*. IEEE Trans. Audio Speech Lang. Process., vol. 19, no. 4, pp. 788–798

[3] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur (2018). *XVectors: Robust DNN Embeddings for Speaker Recognition*. In Proc. 2018 IEEE Int. Conf. Acoust. Speech Signal Process., pp. 5329–5333

[4] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu (2017). *DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN*. ACM Trans. Database Syst., vol. 42, no. 3.

[5] A. Babu, C. Wang, A. Tjandra, K. Lakhotia, Q. Xu, N. Goyal, K. Singh, P. von Platen, Y. Saraf, J. Pino, A. Baevski, A. Conneau, and M. Auli (2022). *XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale*. In Proc. Interspeech 2022, pp. 2278–2282

Students' Acceptance of AI agents: the TAM-AS Model

***Wang Xiaoyun*[1]**

[1]Department of English, School of Foreign Languages, Shandong Jiaotong University, Jinan, China, 3533817706@qq.com

This study addresses the "high-autonomy-low-control" dilemma in language translation education by empowering learners through AI agents that make dynamic decisions. An important language-translation skill is that of translation revision. To support development of translation revision skills for university-level English language students, we designed an AI partner (implemented as a multi-agent system) with three autonomous yet coordinated capabilities: real-time task planning, context-aware error fixes, and cognitive attunement. Evaluation with 120 language-translation students showed our AI partner boosted independent revision skills by 47% while making technology acceptance 23.6% more predictable than when using older tools. The key? Letting AI adapt like a human coach – guiding strategically, without taking over.

Problem statement. Current research in translation pedagogy focuses on passive use of static Computer-Assisted Translation (CAT) tools (Bowker & Fisher, 2010.). However, dynamic AI-based language translation can empower learners' agency beyond CAT effects and greatly increase learning efficiency. To this end, this study proposes the TAM-AS model – extending the Technology Acceptance Model (TAM) (Davis, 1989) with agentic support. It is a multi-agent framework where 3 agents cooperate to help learners develop their translation revision skills. These agents are responsible for: dynamic task planning (real-time optimization), semantic error correction (domain-enhanced contextual intervention), and cognitive attunement (biofeedback-driven thresholds).

Methodology. This three-phase study addressed the underexplored role of AI agent autonomy in translation pedagogy:

- Phase 1 - specifying and implementing 3 cooperating agents in the TAM-AS framework, to carry out the following tasks: dynamic task planning (algorithm-adjusted workflows), semantic error correction (BERT-enhanced feedback (Devlin et al., 2019)), and cognitive attunement (Tobii eye-tracking-driven interventions (Holmqvist et al., 2011)).
- Phase 2 - engaging 120 third-year translation majors from two Chinese universities in an 8-week course, alternating instructor-led CAT sessions with AI agent tasks. The process included tracking their weekly reports, spotting patterns in their translation revising autonomy, and getting biweekly focus group insights to capture the evolving acceptance dynamics.
- Phase 3 - comparing the effects of using the refined multi-agent TAM-AS and using a non-agentic TAM.

Results.

1. AI agents help improve self-regulated learning

AI agents with smart task planning and helpful feedback helped students make 58% more decisions on their own and improved self-correction by 47% ($p < 0.001$). Their feedback helped students notice mistakes and choose better ways to fix them. This made students move from just following steps to thinking and learning more actively. The focus on Chinese-English translation limits generalizability, but initial simulations confirmed similar consistent trends in French-Chinese translation subsamples.

2. Key determinants of technology acceptance when using the TAM-AS model

The study found that perceived translation quality, clear feedback, and simple interface design were key to students' acceptance of AI agents in the TAM-AS model. High-quality output and easy-to-follow feedback boosted acceptance, while complex interfaces reduced it.

3. Methodological distinctiveness

This study moves beyond traditional evaluations of tool effectiveness in developing translation revision skills by proposing a three-part framework: dynamic task planning, semantic error correction, and cognitive attunement. By combining self-regulated learning theory with non-linear analysis, it reveals key patterns in how students accept and use AI tools, offering new insights into the challenges of balancing AI-based automation and student-directed control in education.

4. Educational applications

The use of TAM-AS model in developing translation revision skills supports a three-phase teaching model – tool use, critical revision, and strategic generation – based on cognitive load theory (Sweller, J. 1988). This approach improves the use of technology, lowers students' anxiety about AI, and boosts advanced translation skills (like Cultural Transposition Techniques,Complex Text Restructuring and Creative Compensatory Translation by over three times. It reflects a shift toward deeper integration of technology and cognition in translation education.

5. Overall

Using the TAM-AS model and the multi-agent system described above has improved the technology acceptance level of English language translation students in this study by 49%, their independent revision skills by 47%, and their advanced translation skills by 300+%.

Keywords: AI agents, translation education technology, Technology Acceptance Model (TAM) extension,elf-regulated learning enhancement.

References:

[1] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008
[2] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 4171–4186.
[3] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285.
[4] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. Oxford University Press.
[5] Bowker, L., & Fisher, D. (2010). Computer-aided translation technology: A practical introduction. University of Ottawa Press.

Semantic Landscapes of Dementia: Insights from PubMed and Language Model Embeddings

Anđelka Zečević[1], Bojana Bašaragin[2], Jovana Kovačević[3], Vladimir Kovačević[4], Elma Dervić[5]

[1]Mathematical Institute, Serbian Academy of Sciences and Arts, andjelkaz@mi.sanu.ac.rs
[2]The Institute for Artificial Intelligence Research and Development of Serbia, bojana.basaragin@ivi.ac.rs
[3]Faculty of Mathematics, University of Belgrade & The Institute for Artificial Intelligence Research and Development of Serbia, jovana@matf.bg.ac.rs
[4]The Institute for Artificial Intelligence Research and Development of Serbia, vladimir.kovacevic@ivi.ac.rs
[5]Medical University of Vienna, Center for Medical Data Science & Complexity Science Hub & Supply Chain Intelligence Institute Austria, dervic@csh.ac.at

According to a report by Alzheimer's Disease International (Evans-Lacko, 2024), there are over 55 million people worldwide living with dementia, a health condition used to describe different brain disorders that affect memory, thinking, behaviour, and emotion. Often referred to as a “democratic disease”, dementia knows no social, economic, or ethnic boundaries, imposing a massive burden not only on the lives of individuals and their caregivers, but also on healthcare systems and the global economy. With projections indicating that the number of people affected by dementia will triple by 2050, early detection and interventions are crucial for slowing down the disease’s progression.

In order to better understand the early onset of dementia, our primary research is based on analyzing and comparing several knowledge sources:

large comorbidity, disease-disease network built on the hospital stays, approximately 45 million, for the entire Austrian population (N=8,996,916) spanning from 1997 to 2014 (Dervić, 2024),

gene-disease databases that structure the genetic knowledge about dementia mechanisms, and

scientific papers that summarize research findings and results.

To validate and assess the novelty of observed disease-disease or gene-disease associations in scientific literature, we rely on PubMed, an openly available repository of abstracts in life sciences and biomedical domain, and its search API. By providing keywords and several filtering options that influence the final result list, we are able to draw preliminary conclusions about the importance of the proposed associations.

Given the availability of biomedical large language models (LLMs) fine-tuned or trained on PubMed abstracts, we found it interesting to compare the insights that can be derived directly by querying the PubMed repository with those based on LLMs. In this talk, we will present an experiment focused on 14 dementia risk factors as outlined in the latest Lancet Dementia Commission Report (Livingston, 2024). By using the embeddings of Microsoft’s BiomedBERT[8], formerly known as PubMedBERT (Gu, 2021), and 2D visualization techniques, we will show how different risk factor scenarios, including various dementia causes (Alzheimer's disease, Lewy body dementia, frontotemporal dementia, and vascular dementia), gender and age descriptions, map into the embedding space. Then, we will compare these embedding-based representations with patterns and associations extracted through direct PubMed queries. Finally, we will highlight both overlaps and divergences in how dementia-related knowledge is captured across the two approaches.

[8] https://huggingface.co/microsoft/BiomedNLP-BiomedBERT-base-uncased-abstract-fulltext

Keywords: biomedical LLMs, PubMed, dementia risk factors

References:

[1] Dervić, E., Sorger, J., Yang, L. *et al.* (2024). Unraveling cradle-to-grave disease trajectories from multilayer comorbidity networks. *npj Digit. Med.* 7, 56. https://doi.org/10.1038/s41746-024-01015-w

[2] Evans-Lacko, S., Aguzzoli, E., Read, S., Comas-Herrera, A., Farina, N. (2024). World Alzheimer's Report https://www.alzint.org/u/World-Alzheimer-Report-2024.pdf

[3] Gu, Y., Tinn, R., Cheng, H., Lucas, M. et al. (2021). Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Trans. Comput. Healthcare* 3, 1, Article 2. https://doi.org/10.1145/3458754

[4] Livingston, G., Huntley, J., Liu, K., et al. (2024). Dementia prevention, intervention, and care: 2024 report of the *Lancet* standing Commission. *The Lancet.* Vol. 404 No. 10452P572-628

Harnessing Large Language Models for Medical Lexicon Simplification

Anđelka Zečević[1], Ranka Stanković[2]

[1]Mathematical Institute, Serbian Academy of Sciences and Arts, andjelkaz@mi.sanu.ac.rs
2Faculty of Mining and Geology, University of Belgrade, ranka@rgf.rs

Access to accurate and comprehensible medical information is crucial for fostering health literacy and making informed safe decisions regarding personal health. However, specialized medical lexicons, indispensable resources within the medical domain, inherently rely on complex terminology, acronyms, and nuanced phrasing that render them largely inaccessible to a lay audience lacking specialized training. This linguistic barrier significantly impedes effective communication and knowledge transfer between medical professionals and the general public, including patients and students. In order to address this challenge, our research focuses on the transformation of medical lexicons into a public-facing resource offering simplified, accessible explanations.

Particularly, our research is based on the medical lexicon in Serbian that has undergone a digitization process within this study. Its current digital format is aligned with the Text Encoding Initiative (TEI) standard, comprising individual lexicon entries (labeled as <entry>). Each entry encompasses a medical headword (labeled as <h>), English or Latin equivalents (labeled as <i>), and a detailed, expert-level definition in Serbian. The core research challenge we tackle is the development of a robust, scalable approach for automatically or semi-automatically simplifying these intricate definitions to a level understandable by individuals without a specialized medical background, while preserving absolute factual accuracy and clinical relevance – a non-negotiable requirement in medical information processing.

Analysis of existing medical lexicons, both traditional publications such as the Medical Lexicon by professor Momčilo Babić and colleagues (Babić, 1999) and contemporary online medical dictionaries, confirms the consistent pattern of employing domain-specific jargon and complex sentence structures underscoring the need for dedicated simplification efforts. The existing structure of our digital lexicon corpus, with its clear <entry>, <h>, and <i> elements, provides a vital foundation for automated processing. For each <entry>, we systematically extract the medical term (from <h>) and the complete, original definition text. The English/Latin equivalents found within the <i> tags are also extracted, serving as valuable supplementary context during the simplification process, potentially aiding in the grounding of terms and cross-lingual understanding for the AI model.

The proposed simplification strategy is primarily based on leveraging the advanced capabilities of large language models (LLM), which currently represent the state-of-the-art for complex text transformation tasks like text simplification. For each lexicon entry, we construct a highly tailored prompt designed for interaction with powerful generative models (initially piloting with models such as GPT-4.1, GPT-4.5, and Gemini 2.5-preview). The prompt explicitly instructs the LLM to act as an AI assistant specialized in medical communication for a defined target audience (e.g., "general public") and rewrite the provided medical definition in clear, simple language, minimizing or explaining specialized jargon.

We employ several prompt engineering techniques to guide the LLM effectively. A crucial element is the use of few-shot learning, where the prompt includes a small number of carefully selected examples comprising pairs of original complex definitions and their desired simplified versions. These examples are essential in demonstrating the target style, level of detail, and simplification approach required. Furthermore, the extracted English/Latin equivalents from the original lexicon entry are incorporated into the prompt as additional contextual information available to the LLM, potentially assisting in resolving ambiguities or providing alternative linguistic framing.

Recognizing that fully automated simplification of complex medical content carries inherent risks regarding accuracy. After the AI generates a candidate simplified definition, it undergoes rigorous review by medical experts. The domain experts evaluate each simplified definition not only for its absolute factual correctness and clinical accuracy but also assess its appropriateness in terms of clarity, readability, and comperhension for the defined target audience. This essential hybrid approach combines the scalability and processing power of LLMs with the indispensable domain expertise and critical judgment of human reviewers, thereby mitigating the risks of generating misleading or incorrect medical information.

In this initial pilot study, we are applying the methodology to a representative small subset of lexicon entries to serve as a proof-of-concept. As every single simplified definition generated by the AI is subjected to thorough review and validation by medical professionals, it allows for detailed analysis of the AI's performance, identification of common errors or areas where simplification is particularly challenging requiring further refinement of the prompting strategy. We also plan to implement iterative prompting cycles, where initial AI outputs are evaluated (potentially using traditional metrics like readability scores, analysis of vocabulary frequency, sentence length, or even feedback loops involving other LLMs evaluating simplicity) and the prompt is refined based on detected issues (e.g., the output remaining too technical, being factually inaccurate, or lacking sufficient detail).

The anticipated outcome of this research is a systematically generated dataset of simplified medical definitions for the Serbian language, derived directly from a high-quality expert source. We will analyze the characteristics of this resulting simplified text, comparing its linguistic and structural properties (e.g., sentence length, vocabulary complexity, use of active vs. passive voice) against the original definitions. This analysis will also help identify specific types of medical concepts, terms, or original definition structures that prove most challenging for the AI-driven simplification process, guiding future improvements.

This research contributes to the fields of AI, NLP, Digital Humanities, and Public Health Informatics by demonstrating a practical and robust methodology for leveraging modern LLMs to make specialized knowledge more accessible. The successful application to a Serbian medical lexicon also addresses the need for high-quality NLP resources and applications in less-resourced languages compared to English. Future work includes scaling the methodology to process the entire digitized lexicon, conducting formal user studies with representatives of the target audience to empirically evaluate the perceived clarity and utility of the simplified definitions, and exploring the integration of this simplified lexicon into interactive public health information platforms or patient education tools.

Keywords: text simplification, medical lexion, large language models

References:

[1] Babić, M. (ed.) (1999) Medicinski leksikon, Obeležja, Beograd.
[2] Stanković, R., Šandrih, B., Stijović, R., Krstev, C., Vitas, D., Marković, A. (2019). SASA Dictionary as the Gold Standard for Good Dictionary Examples for Serbian. In Zingano Kuhn T. Correia M. Ferreria J. P. Jansen M. Pereira I. Kallas J. Jakubíček M. Krek S. Kosem, I. and C. Tiberius, editors, Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference, pages 248–269. 1-3.
[3] Zečević, A., Ćulafić, M., Stojković, S. (2024). On Simplification of Discharge Summaries in Serbian: Facing the Challenges. In *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024*, pages 104–108, Torino, Italia. ELRA and ICCL.

Table of Contents: