

На 27.10.2022 г. независимата Агенция за бизнес новини в Югоизточна Европа SeeNews публикува обширен доклад за развитието на изкуствения интелект в България през 2022 г., включително анализ на състоянието и тенденциите. Наблюдава се видимо нарастване на броя на компаниите, които се интересуват от използването на AI за създаване на иновативни решения. Сумарният общ приход на българските разработчици на изкуствен интелект е достигнал рекордните 25,7 милиона през 2021 г., което е скок с над 35% спрямо сумата през 2019 г. Над 70% от фирмите, работещи в областта, са собственост единствено на български лица или компании. Има тенденция към нарастване на броя на зетите специалисти, като през 2020 г. е достигнат рекорден годишен ръст от 25%.

През 2021 г. близо 40% от зетите в сектора са били ангажирани в научноизследователска и развойна дейност спрямо 29% през 2020 г.

ИИКТ беше поканен да представи за доклада описание на дейности и резултати в областта на изкуствения интелект.

През 2022 г. фокусът на описанието е насочен към резултати, свързани с автоматична обработка на българския език.



Bulgaria's IICT institute co-operates with private and public organisations to cover full NLP range in Bulgarian language

The Institute of Information and Communication Technologies (IICT) conducts fundamental and applied research in the field of computer science, information and communication technologies (ICT) and in the development of innovative interdisciplinary applications of these technologies.



Professor Galia Angelova is Director of the Institute.

Prof. Angelova, one the main fields of operation of the Institute is natural language processing. Many private companies also have projects in that field. Do you co-operate with them or is the Institute developing its own projects?

Natural Language Processing (NLP) is one of the pillars of AI. We have been extensively working on the creation of integrated language resources that would include lexical data, corpora and world knowledge, i.e. lexicons, wordnets, ontologies, knowledge graphs, treebanks as well as general or domain-specific annotated corpora. On top of these resources, we have been also developing language processing pipelines that cover tasks from tokenisation to word sense disambiguation, language models and others. These basic processing modules are used in a wide range of applications including machine translation, creation of domain oriented knowledge processing pipelines, semantic document indexing, knowledge extraction, question answering, multi-document summarisation, keyword extraction, smart education and others. As more advanced applications we list speech processing, intelligent support in eHealth via information extraction from patient documents, machine translation for legal documents including semantic processing, abstractive summarization based on neural language models and construction of a Bulgaria-centric knowledge graph. In 2022, IICT developed and delivered to the Bulgarian Union of the Blind a speech synthesizer using a unique neural network model which automatically puts the word stress to ensure naturally sounding synthesized speech in Bulgarian.

IICT works mainly for Bulgarian language but it also supports the multilingual NLP via integration of Bulgarian language resources with resources for other languages like the Open Multilingual Wordnet, the Universal Dependency Initiative and the ParlaMint Project. The technologies developed within the institute are language independent and can be adapted to other languages given the availability of the corresponding training resources.

IICT often co-operates with companies and public organisations. We work together with Sirma AI (Ontotext) on

tasks related to semantic processing of Bulgarian texts. Our NLP team provided to APIS Bulgarian language resources and technologies for semantic indexing, keyword extraction, extraction-based summarization and machine translation for legal documents. We work with Identrics on training of bias neutral neural language models for Bulgarian, classification of texts with respect to potential fake news content and development of models for abstractive summarization. IICT has developed a highly efficient system for indexing and recognition of advertisements in TV streams for the Bulgarian company H-Tech. In 2015, we participated in the development of the National Diabetes Register which was automatically generated from pseudonymised outpatient records submitted to the Bulgarian Health Insurance Fund. Important indicators like values of blood sugar and glycated haemoglobin levels, body mass index, blood pressure etc. were automatically extracted from unstructured text and integrated into the database. Today the Register is hosted and updated by the University Specialised Hospital for Active Treatment of Endocrinology at Medical University Sofia.

IICT is also using AI in relation to the EU's Green Deal and the implementation of its objectives in Bulgaria. How can AI facilitate this process and in what areas is Bulgaria looking to use AI?

AI enables the analysis of Big Data thus increasing our capacity to better understand, monitor, analyse, control, predict and tackle the environmental challenges as well as the performance of industrial systems that have to adapt to sustainable behaviour. It seems that monitoring and optimization of energy consumption is substantial for Bulgaria as well as the integration of renewable energy sources, and AI can help in this respect. Another important area is agriculture where AI provides tools for more efficient use of land, water, pesticides and fertilisers which could mitigate environmental impacts (agricultural applications of AI are significant anyway, to improve efficiency and farming practices).

What projects does IICT have in the pipeline?

IICT performs numerous projects in areas bordering with AI like decision-making, optimisation and intelligent control so I will mention here only a few examples of our activity. We coordinate CLaDA-BG, the Bulgarian national interdisciplinary research e-Infrastructure for Bulgarian language and cultural heritage, part of the EU infrastructures CLARIN and DARIAH. CLaDA-BG is funded by the Ministry of Education and Science and develops a Bulgaria-centric knowledge graph and AI tools for processing big data, old varieties of Bulgarian, 3D images and other related information.

IICT participates in the Horizon Europe Pathfinder project NEMO BMI (2022-2025) which aims to achieve the first assistance-free neuroprosthetics system. The project addresses the restrictions of current neuroprosthetics devices that require regular retraining within controlled environment by development of an auto-adaptive brain-machine interface which translates motor cortex signals into commands to external effectors. Researchers from this team are elaborating a realistic brain model for HPC simulation of human visual system and conscious perception, within a project funded by the Bulgarian National Science Fund.

The starting project SynGRDiT (Synergy for Green Regional Digital Transformation), which will establish the European Digital Innovation Hub Zagore, is another challenging task.