SLAIF
Slovenska tovarna UI
Slovenian AI Factory

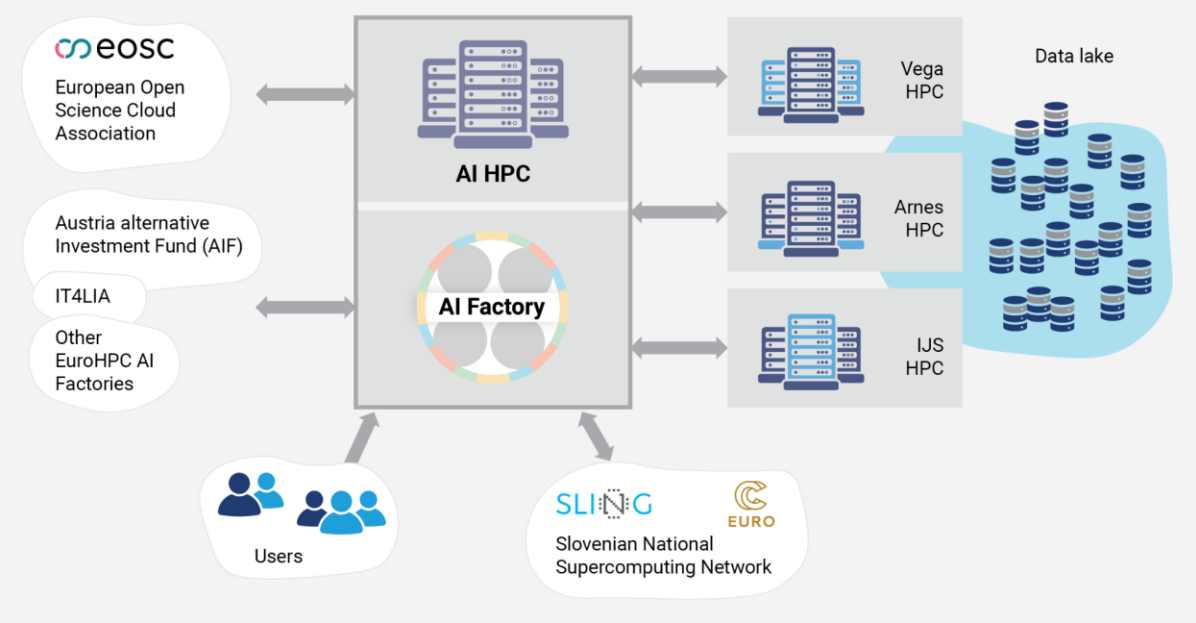
SLAIF
Slovenska tovarna umetne inteligence: kratek uvod

Ekipa SLAIF / prof. dr. Sašo Džeroski, V. Groznik, J. Javoršek

28. maj 2026

Umetna inteligenca in tovarna UI

- Umetna inteligenca je postala povsod prisotno orodje
- Evropa:
  - najprej znanje in razvoj
  - potem regulacija & zakonodaja *(zaupanje)*
  - končno infrastruktura *(suverenost)*
- Tovarna je projekt,
  za UI optimizirani računalnik **šele prihaja**
- Percepcija: veliki jezikovni modeli
  ti pa zahtevajo veliko infrastrukturo:
  - podatkov
  - računalnikov
  - znanja



Kaj so tovarne UI?

- **Kratek odgovor:** Tovarne UI (AI factories) so **infrastruktura** za UI v okviru evropske tehnološke suverenosti in globalne konkurenčnosti:
  - V XX. stoletju so države gradile ceste in električna omrežja
  - V XX1. stoletju morajo graditi infrastrukturo za UI ker zahtevata gradnja in prilagajanje generativnih modelov visokokozmogljive računalnike = HPC

- Vključevati mora **fizično**, **digitalno** (podatke) ter človeško **infrastrukturo** (znanje in storitve).

- Slovenska tovarna UI (SLAIF) nam bo omogočila: da Slovenija ni zgolj uporabnica umetne inteligence, temveč aktivna soustvarjalka evropske tehnološke suverenosti in globalne konkurenčnosti

Zakaj bi imeli tovarno UI?

Tovarne UI znižujejo ovire za uvajanje UI v prakso

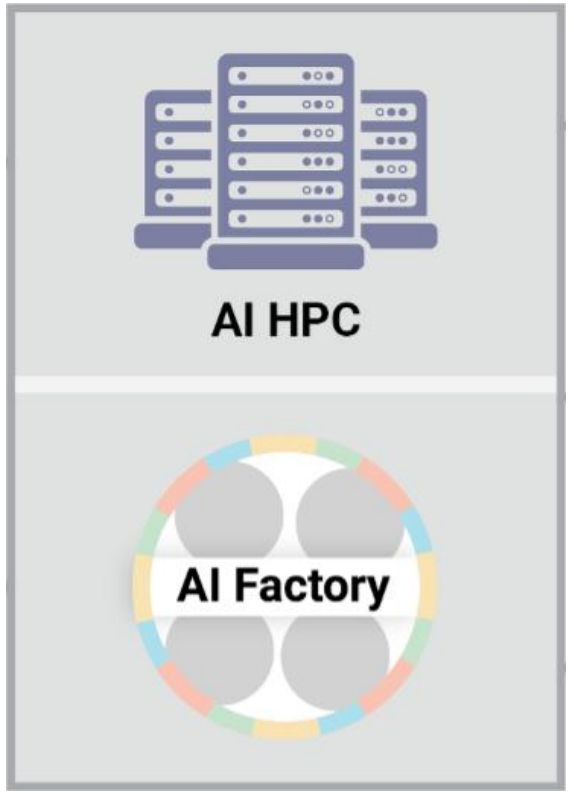
- Ponujajo širok in celovit nabor storitev za industrijo in raziskovalce, ki so potrebne za razvoj evropskih modelov (generativne) UI in njihovo uporabo

Tovarne imajo dve glavni komponenti:

- Superračunalnik, optimiziran za UI
- Storitve in dejavnosti tovarne UI

Tovarne UI združujejo:

- Računsko moč
- Algoritme in delovne tokove
- Podatke in modele
- Veščine, znanje in talente



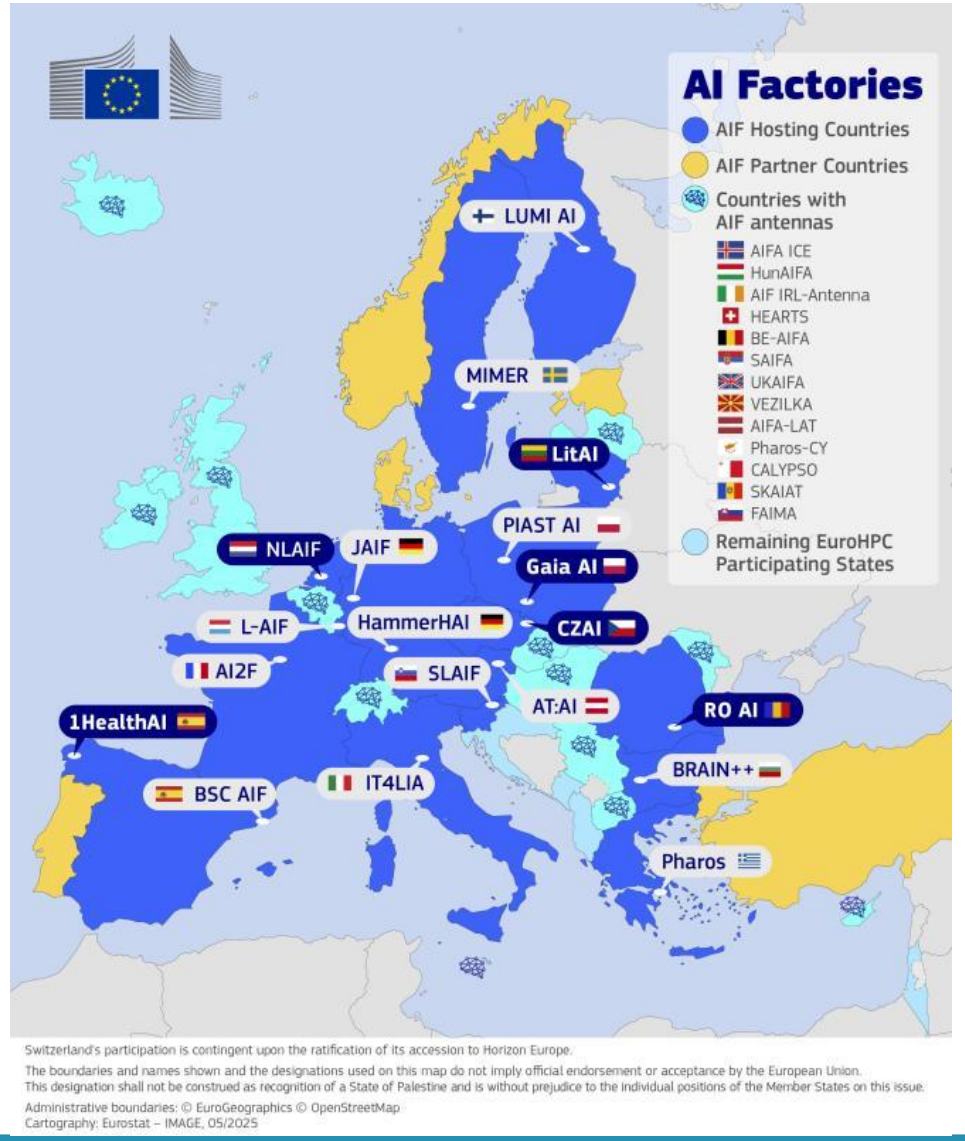
Evropski ekosistem UI v EuroHPC

- EuroHPC in države članice sofinancirajo postavitev **tovarn UI** (s superračunalniki) in **anten** (uporabljajo računalnike tovarn UI):
  - skupaj 19 tovarn UI
  - In 13 anten

    a le 9 tovarn nabavlja in postavlja nove superračunalnike

- Nekaj držav ima deleže v tovarnah UI, ki se vzpostavljajo v drugih državah (npr. Portugalska in Turčija pri BSC AIF)

- Slovenija bo gostila nov superračunalnik, ima pa tudi delež v IT4LIA (10 mio EUR)



Zakaj ima Slovenija lahko tovarno UI?

Čeprav bo imela Evropa številne tovarne UI, jih ne bo imela vsaka država EU. Potrebna je vizija.
Prav tako politična volja in finančna podpora.
In nenazadnje: znanstvena in tehnična podkovanost ter človeški viri:

- **Slovenija ima močno in dolgo delujočo skupnost na področju UI**
- **Slovenija ima močno skupnost na področju visokozmogljivega računalništva:**

  Vega je prvi delujoči sistem EuroHPC
- **Skupnosti odlično sodelujeta** *(kar ni samoumevno)*

Potreben je širši pogled na infrastrukturo

- Infrastruktura ni le strojna oprema (računalniki)
- Podatki so infrastruktura (Slovenija vodi CLARIN, jezikovni podatki)
- UI je infrastruktura (vsekakor za znanost, primer infrastrukturnega centra IJS)

SLAIF – slovenska tovarna UI

Dva projekta, skupna vrednost 135 mio EUR, sofinancerki sta EK in RS (MVZI)

- Superračunalnik, optimiziran za UI, 123.32 mio EUR, 7 let od konca 2027
- Sofinanciran iz programa Digitalna Evropa

Konzorcij projekta sestavljajo:

IZUM Institute of Information Science, Maribor, Slovenia

Jožef Stefan Institute Ljubljana, Slovenia

arnes

- Tovarna UI, storitve in dejavnosti, 11.68 mio EUR, 3 leta od sredine 2025
- Sofinanciran iz programa Obzorje Evropa

Konzorcij projekta sestavljajo:

IZUM Institute of Information Science, Maribor, Slovenia

arnes

University of Maribor

Univerza v Novi Gorici

Gospodarska zbornica Slovenije Chamber of Commerce and Industry of Slovenia

Jožef Stefan Institute Ljubljana, Slovenia

Univerza v Ljubljani University of Ljubljana

FIS Fakulteta za informacijske študije Faculty of information studies

Univerza na Primorskem

TP LJ01

SLAIF – superračunalnik

Kombinacija CPU in GPU enot za podporo tako klasičnemu kot UI-računanju

- CPU (cca 10 petaFLOPS)
- GPU (cca 30 – 1. faza +70 – 2. faza = 100 petaFLOPS) + diskovno polje, hitra povezava

Dve fazi nabave za zagotovitev daljšega obdobja delovanja

- Superračunalnik zagnan do konca leta 2027
- Superračunalnik nadgrajen v letu 2029
- V primerjavi s superračunalnikom VEGA: 3× več moči CPU in 32× več moči GPU

Trenutno se na superračunalniku LEONARDO v Bologni trenirajo:

- Veliki jezikovni model za slovenščino — GaMS (kolegi z UL FRI)
- Prvi vizualno-jezikovni model za slovenščino — SViLa (kolegi z IJS)

takšne naloge bo podpiral novi superračunalnik optimiziran za UI

Generativni model za slovenščino
Generative Model for Slovene
GaMS:

SLAIF – ciljne skupine uporabnikov

Različne skupine uporabnikov

- Gospodarstvo/ industrija
  - Zagonska podjetja
  - Mala in srednja podjetja
  - Velika podjetja
- Javni sektor
  - Zdravstvo
  - Državna uprava
  - Javna podjetja, javni zavodi in agencije
- Raziskovalna sfera (univerze, instituti)

Različne stopnje UI-pripravljenosti

SLAIF – aplikacije in prednostna področja

Zeleni prehod

- precizno kmetijstvo
- okoljsko spremljanje
- optimizacija energije
- pametna proizvodnja

Zdravje in biotehnologija

- personalizirana medicina
- podpora odločanju
- medicinsko slikanje
- genomika in diagnostika
- federativno učenje za analizo zdravstvenih podatkov

Digitalna družba

- jezikovne tehnologije
- mediji in kreativnost
- javni sektor
- izobraževanje

UI za znanost

- podpora raziskovalnemu procesu
- znanosti o življenju
- znanosti o materialih
- digitalna humanistika

1 Zeleni prehod

- Precizno kmetijstvo
  - spremljanje poljščin
  - napovedovanje letine
  - zaznavanje škodljivcev
- Okolje
  - spremljanje sprememb
  - biodiverziteta
  - napovedovanje nesreč
  - upravljanje z vodo
- Optimizacija energije
  - pametna omrežja (smart grids)
  - integracija obnovljivih virov
  - napovedovanje potreb
  - samodejno upravljanje napetosti
- Pametna proizvodnja
  - vizualni nadzor kvalitete
  - napoved potreb po vzdrževanju
  - digitalni dvojčki
  - podpora krožni ekonomiji

2 Biotehnologija in medicina

- Personalizirana medicina
  - modeli z genomskim in kliničnimi podatki
  - prilagojene terapije
  - nasveti za življenjski slog
- Medicinsko slikanje in simulacija
  - optimizacija odločitev
  - izboljašanje inovacij
  - napovedovanje razvoja bolezni
- Genomika in diagnostika
  - analiza biosgnalov (ECGs, EEGs)
  - medicinskih slik (MRI, CT)
  - integracija podatov za osebno in razložljivo diagnostiko
- Odkrivanje zdravil
  - napovedovanje lastnosti učinkovin
  - razvoj usmerjenih zdravil
  - zdravljenje redkih bolezni
  - načrtovanje sinteze

3 Digitalna družba

- Jezikoslovne tehnologije
  - veliki jezikovni modeli za slovenščino
  - napredno samodejno prepoznavanje govora (ASR)
  - prevajanje in povzemanje
  - anonimizacija
- Mediji in ustvarjalnost
  - zaznavanje dezinformacij
  - preverjanje dejstev, novinarstvo
  - oglasi, rendering
- Javni sektor
  - občanska znanost
  - pomoč pri upravnih postopkih
  - razvoj zakonodaje in pravilnikov
  - pametna mobilnost
- Vzgoja in izobraževanje
  - personalizirani inštruktorji
  - avtomatizirano ocenjevanje
  - prilagodljivo oz. vodeno učenje

Umetna inteligenca za znanost

- Podpora raziskavam
  - agenti z velikimi modeli
  - simbolična regresija
  - samodejno postavljanje hipotez
- Znanosti o življenju
  - virtualno presejevanje
  - high-throughput simulations
  - genomika
- Znanost o materialih
  - novi materiali za energijo
  - Katalizatorji, lastnosti materialov
  - trajnostna proizvodnja
- Digitalna humanistika
  - veliki modeli za jezikoslovje
  - zgodovinopisje
  - podpora pravu
  - raziskave na področju kulture in umetnosti

SLAIF – storitve in dejavnosti

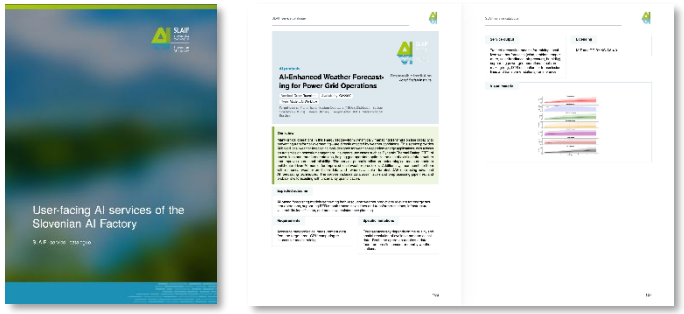
Dostop do računskih in podatkovnih zmogljivosti

Splošni delotoki za učenje in prilagajanje modelov

- Veliki jezikovni modeli
- Veliki govorni modeli
- Vizualno-jezikovni modeli
- Večmodalni modeli

Domensko-prilagojeni delotoki (in modeli) za različne vertikale

- Zeleni prehod
- Zdravje in botehnologija
- Digitalna družba
- UI za znanost

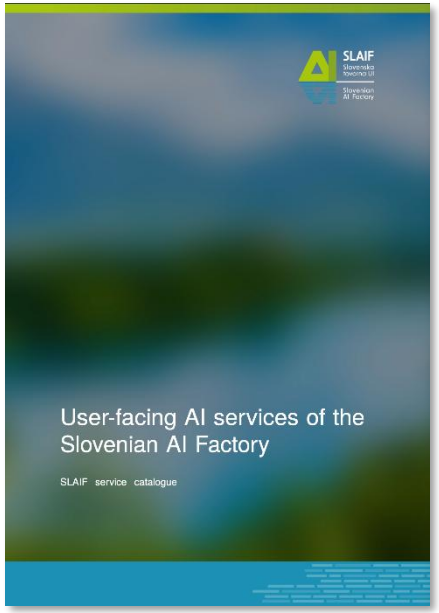


Izobraževanje in pridobivanje veščin UI

SLAIF – katalog storitev

Več vrst storitev

- podpora uporabnikom
- infrastruktura in podatki
- razvoj in uvajanje UI
- produkti UI



Contents

User support 9
Business Processes Review and Automation Using AI 11
Support at Distributed Training of Large-Scale AI Models 12
Support in Model Optimisation: Pruning, Quantisation, and Distillation 14
Support in Proof-of-Concept (PoC) Application Development 16
Technical User Support 18
Infrastructure and data-related services 19
Cloud Platform Service (OpenStack PaaS) 21
Container Image Registry 22
Container Platform Service (Kubernetes) 23
Data Access Services 24
Data Analytics Services 26
Semantic Data Discovery over the SLAIF Catalogue 28
Data Governance Services 30
Data Management / Sharing Services – Connectors for Trusted Sharing 31
Data preparation services 33
Data Processing Services 35
Exploration and Investigation of Disaster Events through the Slovenian EO Data Cube 37
Facebook PE-Video with Slovenian Annotations 39
FAIR-Compliant Metadata Management and Discovery Service 40
Fast Parallel Storage 42
Foundation-Model-Assisted Dataset Bootstrapping from Uncurated Visual Data (Auto-Annotation) 43
HPC Accelerated Compute Service 45
HPC Compute Service 47
Mid- and Long-Term Storage 48
Mid- and Long-Term Remote Access Storage 49
Monitoring Services 50
Orchestration Services 51
Synthetic Data Services 52
Synthetic Genetic Material Data Generation 54
Transformation and Processing Pipeline for Genetic Material Data 55
User Software Repository 57
AI development and deployment 58
Foundation Models for Human Sensing with Support for Fine-Tuning 60
Adaptation of a Large Language Model for Domain-Specific Question Answering 62
Adaptation of a Large Language Model for Domain-Specific Summarization 63
Adaptation of a Large Language Model for a Specific Task using Parameter Efficient Fine-Tuning 64
Adaptation of a Large Language Model for Sentiment and Emotion Classification using Parameter Efficient Fine-Tuning 65
Adaptation of a Large Language Model to a Specific Domain using Instruction Following 66
AI-Based Image Analysis for Industrial Anomaly Detection and Quality Control 67
AI-Based Microscopy Image Analysis 69
Automated Construction of Domain-Specific Knowledge Graphs 71
Change Detection and Description from Remotely Sensed Data 73
Embedding Selection – Intelligent Embedding Selection for Scalable AI Systems 75
ML Pipeline Exploiting Known Attribute Value Uncertainties 77
Model Evaluation for the Local Context Using Linguistic and Cultural Competence Measures 78
Per-Prompt LLM Selection – Automatically Choosing the Best Open Access LLM 80
Preference Optimization Pipeline and LLM-Based Machine Translator for English and Slovene Pairs 82
Satellite and Aerial Image Analysis 83
Temporal, Multi-Modal Foundation Models from Remotely Sensed Information 85
Visual and Interactive AI Workflows (OrangeWeb) 87
Visual and Interactive AI Workflows (OrangeWeb) for Green Transition 88
Visual and Interactive AI Workflows (OrangeWeb) for Health and Biotechnology 89
AI products 90
Access to a Multimodal Language/Sensor Model for Data Analysis 92
Agentic RAG for Administration 94
AI-Assisted Workflow for Grading Scanned Handwritten STEM Exams 96
AI-Based Human Pose Estimation and Kinematic Analysis 98
AI-Based Predictive Analytics for Strategic Resource and Trend Forecasting 100
AI-Based Visual Defect Detection and Quality Assessment for Industrial Production 102
AI-Driven Predictive Maintenance and Performance Optimization for Energy Systems 104
AI-Enhanced Weather Forecasting for Power Grid Operations 106
AI for Health and Biotechnology – AI-Assisted Lab Blood Test Interpretation 108
AI Framework for ECG Analysis 110
AI-Powered Intelligent Document Analysis and Semantic Retrieval 112
AIDA – Agricultural Intelligence and Decision Assistant 114
Automated Classification and Feature Extraction for Experimental Materials Data 116
Crop Growth Monitoring and Analytics Services 118
Domain-Agnostic Object Counting, Detection and Tracking in Images 120
Energy Distribution Optimisation 122
FoodyLLM – AI Copilot for Food Nutrition Intelligence 123
Irradiance Prediction Service for Energy Modelling 125
Multi-Label Classification for Microbial Data 127
Multilingual Idiom Semantic Search and Translation System 128
Observable-Cue Mental Health Risk Prediction & Screening Toolbox 130
Open Model for Slovenian Acoustic Prosody Annotation 132
Pre- and Post-Surgery Primary Brain Tumour Volume Estimation 133

SLAIF – katalog storitev

AI development and deployment

Services for adapting, fine-tuning, integrating, scaling, and operationalising AI workflows and models for real use cases.

AI development and deployment

Adaptation of a Large Language Model for Domain-Specific Question Answering

Responsible institution
University of Ljubljana

Vertical Cross-domain | Availability Q3/2027 | Type Pipeline

Target users: Industry, Large Enterprises, Scientific Institutions, Public Institutions

Overview

This pipeline enables the adaptation of a large language model to answer domain-specific questions accurately and efficiently. It is designed to support expert systems, knowledge bases, or FAQ systems in specialized fields. The prerequisites are a sufficient amount of domain-specific data and domain know-how. The pipeline first helps users in preparing data for question answering based on the domain data and provides code and know-how for synthetically expanding the training data based on the provided domain data. Next, it provides the know-how and code for fine-tuning the model with the created instruction-tuning dataset. Finally, it provides code for efficient inference using the adapted model. The pipeline leverages the Slovene LLM GaMS and similar open-source models, tools for generating domain-specific question-answer pairs using NeMo Curator in combination with the GaMS-Instruct approach, LoRA or QLoRA-based supervised fine-tuning using Hugging Face and NeMo, and vLLM for fast domain-specific question answering inference.

Expected outcome

A step-by-step guide for domain-specific question-answering training, along with scripts for data generation, model training, and inference.

Requirements

Requires several NVIDIA GPUs.

Specific limitations

Service output

Pipeline for question answering adaptation, including data generation scripts, model training scripts, and inference code.

Licensing

MIT

AI development and deployment

Adaptation of a Large Language Model for Domain-Specific Summarization

Responsible institution
University of Ljubljana

Vertical Cross-domain | Availability Q3/2027 | Type Pipeline

Target users: Industry, Large Enterprises, Scientific Institutions, Public Institutions

Overview

This pipeline enables the adaptation of a large language model (LLM) to generate high-quality, domain-specific summaries for technical or scientific documents, reports, research papers, or industry-specific content. It is designed to enhance the accuracy and relevance of summaries within a specific domain. The pipeline first helps users in preparing data for summarization (OCR of PDF documents) and converting it into a training dataset. Next, it provides the know-how and code for fine-tuning the model with the created dataset. Finally, it provides code for efficient inference using the adapted model. The pipeline leverages the Slovene LLM GaMS (for Slovene-language summarization) and other open VLMs (e.g., Gemma), open-source OCR libraries such as Marker or small open-source VLMs such as Nanonets, supervised fine-tuning or preference tuning using GRPO with LoRA or QLoRA-based fine-tuning, and vLLM for efficient, low-latency inference of domain-specific summarization tasks.

Expected outcome

A step-by-step guide for domain-specific summarization adaptation, along with scripts for data preparation, model training, and inference.

Requirements

Requires several NVIDIA GPUs.

Specific limitations

Service output

Pipeline for summarization adaptation, including data preparation scripts, model training scripts, and inference code.

Licensing

MIT

SLAIF – izobraževanje in usposabljanje

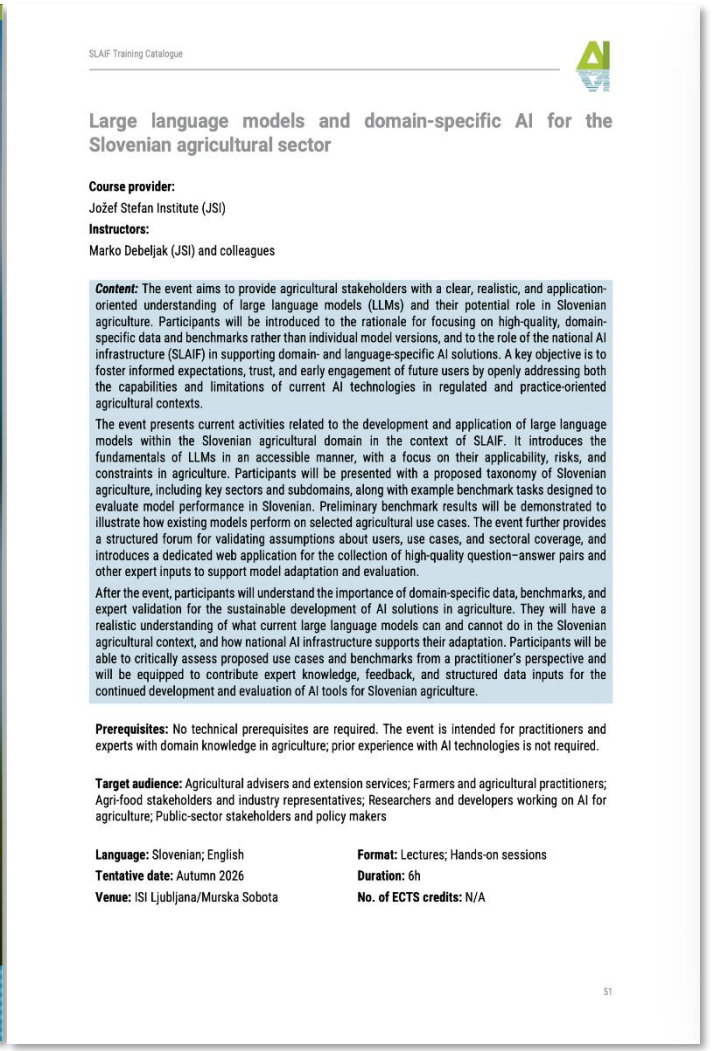
Različne vrste izobraževanj

- Delavnice
- Webinarji
- Izobraževanja z mikro-certifikati

Različne tematike izobraževanj

- UI metode
- Visokozmogljivo računalništvo
- Področja uporabe UI

Različne ciljne skupine



SLAIF Training Catalogue

Large language models and domain-specific AI for the Slovenian agricultural sector

Course provider:
Jožef Stefan Institute (JSI)
Instructors:
Marko Debeljak (JSI) and colleagues

Content: The event aims to provide agricultural stakeholders with a clear, realistic, and application-oriented understanding of large language models (LLMs) and their potential role in Slovenian agriculture. Participants will be introduced to the rationale for focusing on high-quality, domain-specific data and benchmarks rather than individual model versions, and to the role of the national AI infrastructure (SLAIF) in supporting domain- and language-specific AI solutions. A key objective is to foster informed expectations, trust, and early engagement of future users by openly addressing both the capabilities and limitations of current AI technologies in regulated and practice-oriented agricultural contexts.

The event presents current activities related to the development and application of large language models within the Slovenian agricultural domain in the context of SLAIF. It introduces the fundamentals of LLMs in an accessible manner, with a focus on their applicability, risks, and constraints in agriculture. Participants will be presented with a proposed taxonomy of Slovenian agriculture, including key sectors and subdomains, along with example benchmark tasks designed to evaluate model performance in Slovenian. Preliminary benchmark results will be demonstrated to illustrate how existing models perform on selected agricultural use cases. The event further provides a structured forum for validating assumptions about users, use cases, and sectoral coverage, and introduces a dedicated web application for the collection of high-quality question–answer pairs and other expert inputs to support model adaptation and evaluation.

After the event, participants will understand the importance of domain-specific data, benchmarks, and expert validation for the sustainable development of AI solutions in agriculture. They will have a realistic understanding of what current large language models can and cannot do in the Slovenian agricultural context, and how national AI infrastructure supports their adaptation. Participants will be able to critically assess proposed use cases and benchmarks from a practitioner's perspective and will be equipped to contribute expert knowledge, feedback, and structured data inputs for the continued development and evaluation of AI tools for Slovenian agriculture.

Prerequisites: No technical prerequisites are required. The event is intended for practitioners and experts with domain knowledge in agriculture; prior experience with AI technologies is not required.

Target audience: Agricultural advisers and extension services; Farmers and agricultural practitioners; Agri-food stakeholders and industry representatives; Researchers and developers working on AI for agriculture; Public-sector stakeholders and policy makers

Language: Slovenian; English
Tentative date: Autumn 2026
Venue: ISI Ljubljana/Murska Sobota
Format: Lectures; Hands-on sessions
Duration: 6h
No. of ECTS credits: N/A

51

SLAIF – izobraževanje in usposabljanje

88
načrtovanih izobraževalnih aktivnosti

59
delavnic in praktičnih usposabljanj

16
webinarjev

13
mikrodokazil, usklajenih z EQF

Safe, responsible and compliant AI for SMEs: From generative AI to the EU AI Act

Course provider:
Jožef Stefan Institute (JSI)
Instructors:
Maja Škrjanc (JSI), Mitja Jermol (IRCAI)

Content: This 2-hour training is designed for SMEs, small and micro companies that are introducing—or considering the introduction of—artificial intelligence, with a particular focus on generative AI. Participants receive a structured overview of core AI concepts (including the distinction between data-driven methods and GenAI, and the role of context, memory and reasoning), alongside a business-oriented discussion of typical risks in practice (e.g., confidentiality and data leakage, GenAI misuse, deepfakes and brand/identity fraud).

FAIR data management for artificial intelligence

Course provider:
Jožef Stefan Institute (JSI)
Instructors:
Panče Panov (JSI)

Learning objectives: 1) Understand AI data assets across the lifecycle: how datasets, labels, dataset splits, features, and evaluation artifacts evolve from collection to reuse; 2) Apply the FAIR principles to AI work: make data and outputs easier to find, access, combine, and reuse (for teams and future projects); 3. Create an actionable DMP for AI projects: a lightweight plan that supports reproducibility, handover, and compliance; and 4) Handle constraints responsibly: recognize sensitive data, ethical considerations, access limitations, and industry vs research expectations.

Efficient LLMOps, hosting, and model quantization

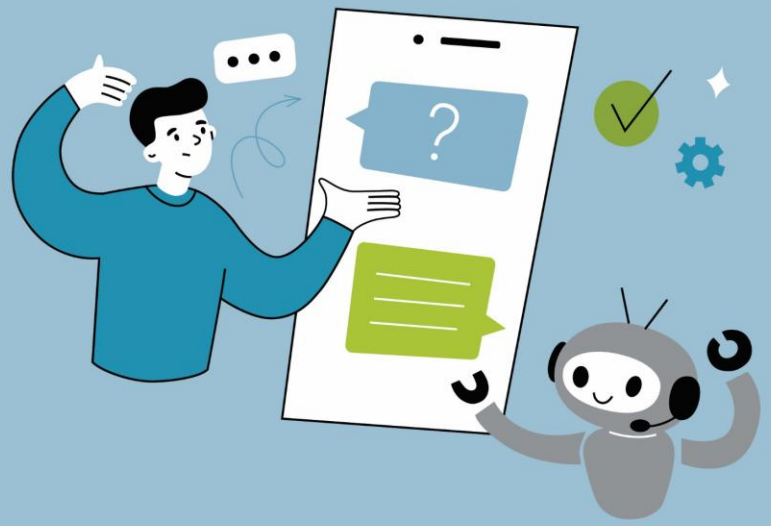
Course provider:
Jožef Stefan Institute (JSI)
Instructors:
Boshko Koloski (JSI), Matej Martinc (JSI), Usama Derebashi (JSI), Nikola Marić (JSI), Sašo Džeroski (JSI)

Content: This course covers the operational aspects of large language models (LLMOps), efficient hosting, and model quantization. LLMOps represents an extension of MLOps for generative models; it includes practices for managing the entire lifecycle of LLMs—from selection and customization to deployment and continuous monitoring. Participants learn the differences between MLOps and LLMOps and why generative models require additional procedures, such as robust data pipelines, version control, and flexible resource orchestration.

Large language models and domain-specific AI for the Slovenian agricultural sector

Course provider:
Jožef Stefan Institute (JSI)
Instructors:
Marko Debeljak (JSI) and colleagues

Content: The event aims to provide agricultural stakeholders with a clear, realistic, and application-oriented understanding of large language models (LLMs) and their potential role in Slovenian agriculture. Participants will be introduced to the rationale for focusing on high-quality, domain-specific data and benchmarks rather than individual model versions, and to the role of the national AI infrastructure (SLAIF) in supporting domain- and language-specific AI solutions. A key objective is to foster informed expectations, trust, and early engagement of future users by openly addressing both the capabilities and limitations of current AI technologies in regulated and practice-oriented agricultural contexts.



Kaj je končni cilj tovarne?

- **Za vas:** razvita storitev, nova tehnologija, pilotni primer
- **Za nas:** dosežek
- **Ne pa:**
  - prostor za komercialno uporabo
  - dolgotrajna shramba za podatke

Offboarding
poročilo
dokumentacija
promocija

Vendar pa tudi

- **Forum** (in prostor) za sodelovanje
- **Posredovanje** med ponudniki in uporabniki
- Možnost za **testiranje** in **skaliranje**
- **Mednarodne** povezave

Register
ekspertnih
podjetij in
izvajalcev

Hvala za udeležbo! Vprašanja?

Thank you for attending! ***Questions?***

SLAIF
Slovenska tovarna UI
Slovenian AI Factory

IZUM
Institute of Information Science, Maribor, Slovenia

arnes

University of Maribor

Univerza v Novi Gorici
Scientia Vincès 1995

Gospodarska zbornica Slovenije
Chamber of Commerce and Industry of Slovenia

Jožef Stefan Institute
Ljubljana, Slovenia

UNIVERZA V LJUBLJANI
University of Ljubljana

FIS
Fakulteta za informacijske študije
Faculty of information studies

Univerza na Primorskem
Università del Litorale

TP LJ01

Financerja / Financed by:

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA VISOKO ŠOLSTVO, ZNANOST IN INOVACIJE

EuroHPC
Joint Undertaking

Projekt SLAIF: Slovenska tovarna umetne inteligence je finančno podprlo Ministrstvo za visoko šolstvo, znanost in inovacije. Projekt je bil na razpisu skupnega podjetja EuroHPC izbran za financiranje v okviru programov Obzorje Evropa ter Digitalna Evropa.

SLAIF: Slovenian AI Factory has been funded by the Ministry of Higher Education, Science and Innovation of Republic of Slovenia. At a call by EuroHPC JU, the project has received a positive funding decision under Horizon Europe and Digital Europe Programmes.