SLAIF
Slovenska tovarna UI
Slovenian AI Factory

SLAIF — Slovenska tovarna UI

Umetna inteligenca za javni sektor:

priložnosti, orodja in prvi koraki

Ljubljana, 18. maj 2026

Prof. Sašo Džeroski – tehnični vodja HE projekta SLAIF

ZAKAJ DANES VSI GOVORIJO O UMETNI INTELIGENCI?

Ker lahko UI danes rešuje številne vsakodnevne probleme

- UI je danes vključena v številne izdelke in storitve (iskanje, prevajanje, podpora uporabnikom, …)
- Ključ do tega je uporaba splošnonamenskih temeljnih modelov

Kaj so temeljni modeli (FM)?

- Temeljni modeli so veliki modeli umetne inteligence, naučeni na obsežnih splošnih podatkih
- Kasneje jih je mogoče prilagoditi (uglasiti) z uporabo manjših količin podatkov za reševanje specifičnih nadaljnjih nalog

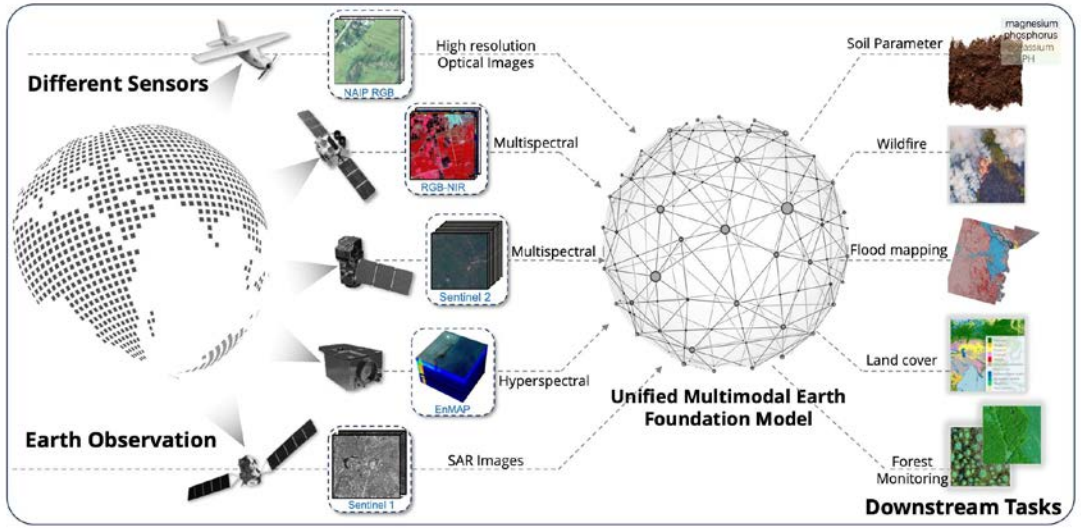
Veliki jezikovni modeli (LLM)

- LLM-ji so temeljni modeli, naučeni na besedilih
- Običajno se učijo z napovedovanjem naslednje besede
- Lahko se uporabljajo za generiranje besedila (generativna umetna inteligenca)
- Nadaljnje naloge vključujejo odgovarjanje na vprašanja, povzemanje, …

UČENJE IN PRILAGAJANJE FM MODELOV ZAHTEVA VELIKO RAČUNSKE MOČI

Poleg besedila lahko temeljni modeli obdelujejo tudi druge podatke, npr. različne vrste slik

Temeljni modeli so velike nevronske mreže
Z milijardami parametrov (moč povezav med nevroni)



- **Učenje in prilagajanje LLM/FM modelov zahteva veliko računske moči**
- LLM/FM modeli se učijo iz ogromnih količin podatkov, kar zahteva zelo veliko računsko moč (več tednov ali mesecev izvajanja na GPU-jih)
- Tudi prilagajanje modela (fine-tuning), pri katerem FM prilagodimo specifični domenski nalogi, še vedno zahteva znatne računske vire

PRISTOP EVROPSKE UNIJE K UMETNI INTELIGENCI: pospešitev sprejemanja in spodbujanje inovacij

Regulativa: Akt o umetni inteligenci (AI Act)

- Varnost
- Zaupanje

Infrastruktura

- Tovarne umetne inteligence
- Omogočanje razvoja (generativnih) modelov umetne inteligence

Inovacije

- Apply AI (Uporabi UI): pospeševanje uvajanja UI v industriji in javnih storitvah
- Hkrati tudi podpora raziskavam (npr. AI in Science / RAISE)

KAJ SO TOVARNE UI?

Kratek odgovor: ***Tovarne UI (AI factories) so infrastruktura za UI***

- V 20. stoletju so države gradile ceste in električna omrežja
- V 21. stoletju morajo graditi tudi infrastrukturo za UI (ker zahtevata gradnja in prilagajanje generativnih modelov visokokozmogljive računalnike=HPC)
- To je vprašanje tehnološke suverenosti in globalne konkurenčnosti

Slovenska tovarna UI (SLAIF) nam bo omogočila,

da Slovenija ni zgolj uporabnica umetne inteligence,

temveč **aktivna soustvarjalka evropske tehnološke suverenosti**

in globalne konkurenčnosti

KAJ SO TOVARNE UI?

Tovarne UI znižujejo ovire za uvajanje UI v prakso

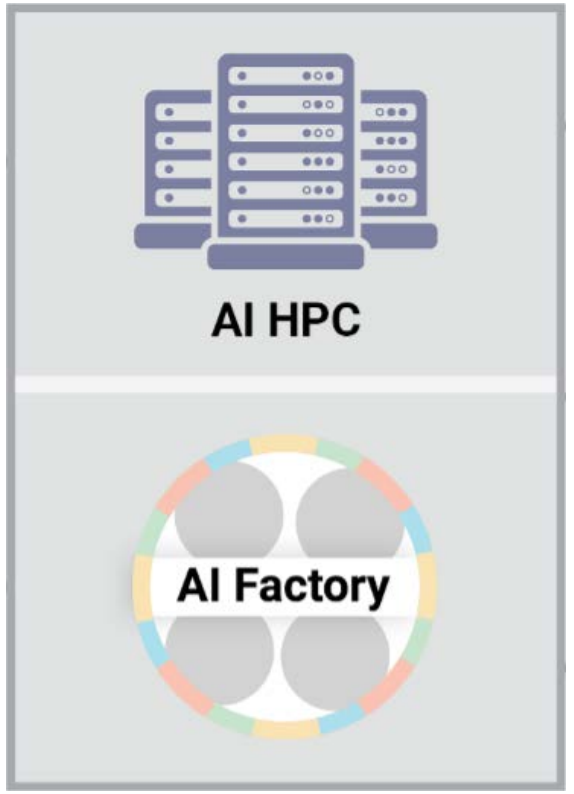
- Ponujajo širok in celovit nabor storitev za industrijo in raziskovalce, ki so potrebne za razvoj evropskih modelov (generativne) UI in njihovo uporabo

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

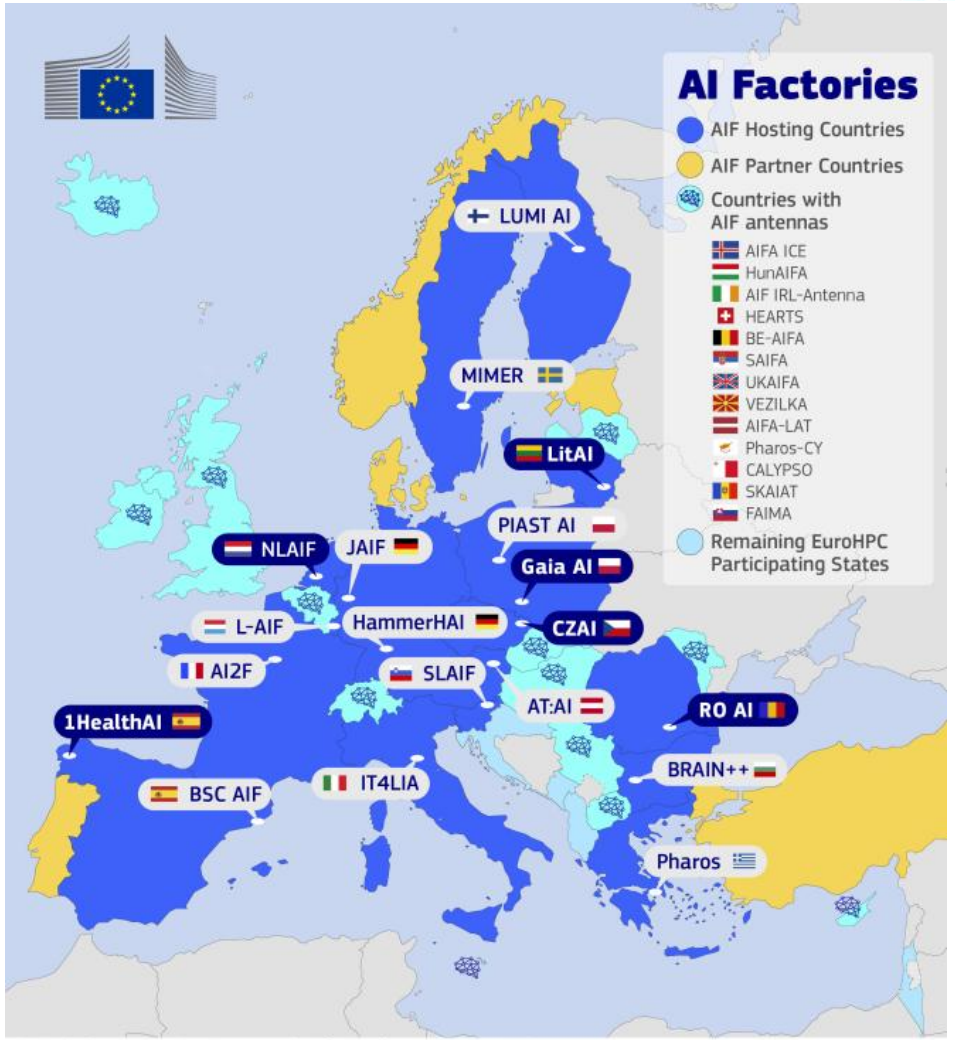


EVROPSKI EKOSISTEM TOVARN UI

- EuroHPC in države članice sofinancirajo postavitev **tovarn UI** (s superračunalniki) in **anten** (uporabljajo računalnike tovarn UI)
- Skupaj 19 tovarn UI
- Skupaj 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 LAHKO SLOVENIJA 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 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
  - Sofinanciran iz programa Digital Europe
  - 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
  - Sofinanciran iz programa Horizon Europe
  - 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)

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 – STORITVE IN DEJAVNOSTI

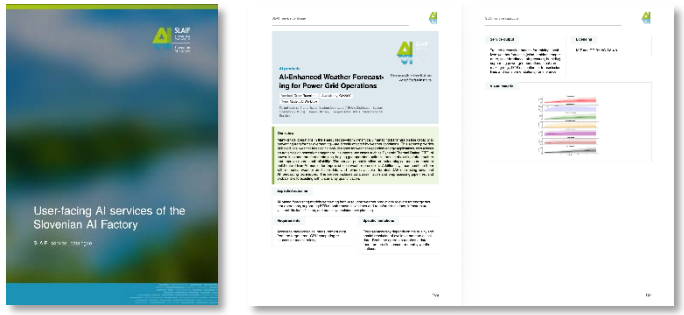
Dostop do računskih in podatkovnih zmogljivosti

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

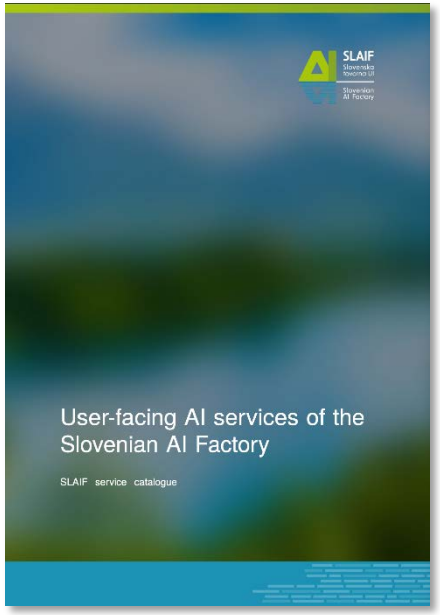


Izobraževanje in pridobivanje veščin UI

SLAIF – STORITVE

Več vrst storitev

- Podpora uporabnikom
- Infrastruktura in podatki
- Razvoj in uvajanje UI
- Produkti UI



Contents

SLAIF – STORITVE

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

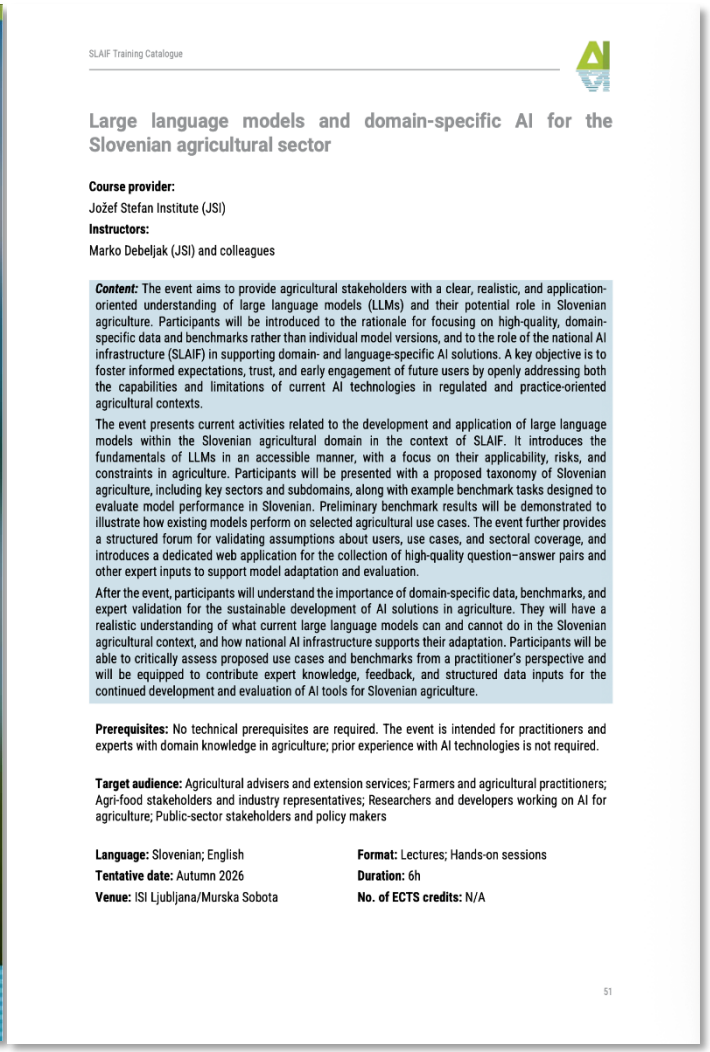
Različne vrste izobraževanj

- 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.

Hvala za udeležbo!

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA VISOKO ŠOLSTVO, ZNANOST IN INOVACIJE

SLAIF: Slovenska tovarna umetne inteligence je prejela financiranje Skupnega podjetja za evropsko visokozmogljivo računalništvo v okviru sporazuma o nepovratnih sredstvih št. 101254461. Projekt financirata Ministrstvo za visoko šolstvo, znanost in inovacije Republike Slovenije ter program Horizon Europe Evropske unije.

SLAIF: Slovenian AI Factory has received funding by the European High Performance Computing Joint Undertaking under the grant agreement no. 101254461. The project is funded by the Ministry of Higher Education, Science and Innovation of Republic of Slovenia and by Horizon Europe Programme of the European Union.