

HORIZON-WIDERA-2023-ACCESS-02

# ChatMED

Grant ID: 101159214

**Bridging Research Institutions to Catalyze Generative AI Adoption by the Health Sector in the Widening Countries**



"Ss. Cyril and Methodius" University in Skopje  
FACULTY OF COMPUTER  
SCIENCE AND ENGINEERING



Jožef  
Stefan  
Institute



Univerzitet u Nišu  
MEDICINSKI FAKULTET



---

## Deliverable D4.1

# Report on Training, STSEs, and Summer School Activities in Period 1

---

<b>Deliverable D4.1 - Report on Training, STSEs, and Summer School Activities in Period 1</b>	
<b>Date:</b>	<b>August 29, 2025</b>
<b>Lead Beneficiary:</b>	<b>JSI</b>
<b>Type:</b>	<b>R</b>
<b>Author(s):</b>	<b>Monika Simjanoska Misheva, Primož Kocuvan, Matjaž Gams</b>
<b>Dissemination level:</b>	<b>PU</b>
<b>Version:</b>	<b>1.0</b>
<b>Abstract:</b> Report includes first year training, STSE and summer school activities.	
<b>Disclaimer:</b> The project is funded by the European Union under Horizon Europe Widening participation and spreading excellence programme, grant ID: 101159214. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.	

<b>HISTORY OF CHANGES</b>			
<b>Date</b>	<b>Issue</b>	<b>Author</b>	<b>Summary of main changes</b>
10.7.2025	1.0	Monika S. Misheva	Wrote distinct reports on the training sessions, STSE, and summer school.
15.7.2025	2.0	Primož Kocuvan	Completed the report on training sessions, STSE, and summer school.
26.7.2025	3.0	Primož Kocuvan	Updates on the report, proof-read.



## Table of Contents

Training Report: Year 1 Project Activities – Comprehensive AI Workshops.....	4
Report on training sessions.....	9
Training Session 1: Foundations of Generative AI.....	9
Training Session 2: AI for Neurological Applications.....	17
Results from the Training Feedback Form.....	27
Report on STSE.....	28
STSE Report for Topic 1.....	28
STSE Report for Topic 2.....	30
STSE Report for Topic 3.....	32
STSE Report for Topic 4.....	34
Report on summer school.....	36
The Agenda design.....	40
Daily Implementation and Outcomes.....	40
Summary.....	43
Technical and AI-Related Skills:.....	44
Medical and Clinical Skills:.....	44
Regulatory and Compliance Skills:.....	44
Collaborative and Proposal Development Skills:.....	45
Soft and Organizational Skills:.....	45

# Training Curriculum Report for Year 1

## Training Report: Year 1 Project Activities – Comprehensive AI Workshops

### **Host Institution:**

Faculty of Computer Science and Engineering (FCSE), Skopje, N. Macedonia

### **Training session 1 (number of participants):**

25 April – 26 April 2025 (36 participants confirmed + temporary listeners from the conference)

### **Training session 2 (number of participants):**

9 July 2025 (22 participants)

During the first year of project implementation, the Faculty of Computer Science and Engineering (FCSE), acting as the project coordinator under Deliverable D2.4 “Detailed Training Curriculum and Talent Mobility Programs,” organized two comprehensive training sessions at the Widening coordinator’s premises in conjunction with the 22nd International Conference on Informatics and Information Technologies (CiiT) in Strumica, North Macedonia. The initial two-day event took place on April 25–26, 2025 as planned; the second session, originally intended to follow immediately, was postponed to July 9, 2025 in response to overwhelming industry interest and to allow additional time for networking and collaboration. Both sessions combined theoretical instruction with practical, hands-on workshops led by the ChatMED team, each designed to bridge the knowledge gap in Generative AI applications within healthcare, especially neurology and to equip researchers and practitioners with essential skills and insights.

Prof. Dimitar Trajanov delivered an overview of current trends in Generative AI and its transformative potential in medical contexts. Asst. Ana Todorovska led a practical session on prompt engineering, emphasizing strategies to refine and enhance the performance of large language models. Asst. Jovana Dobrova introduced Retrieval-Augmented Generation frameworks, demonstrating how they bolster information retrieval and contextual reasoning in clinical scenarios. Prof. Katarina Trojancanec Dineva focused on predictive modeling applications in neurology for improved diagnosis and patient care, while Prof. Ivan Kitanovski explored the integration of AI with medical imaging, specifically MRI and CT data interpretation to advance diagnostic accuracy. Finally, Prof. Riste Stojanov presented domain-driven software design principles for creating customized, interoperable solutions tailored to the unique challenges of neurological healthcare. In more detail is presented in Table 1. Follow-up blog posts are planned

to elaborate on the tools, methodologies and clinical implications discussed, ensuring that the broader healthcare and AI communities can leverage the event's outcomes.

The core objectives of the Year 1 training program, as outlined in Deliverable D2.4, were to:

- Introduce participants to the foundational principles of Generative AI.
- Provide hands-on experience with popular AI frameworks.
- Trigger effective academic and industry collaborations in AI research.
- Address the customization of software solutions to meet neurological needs and challenges.
- Enhance expertise in AI applications within healthcare and neurology.

These objectives are directly aligned with the broader goals of the ChatMED project, aiming to establish a comprehensive Generative AI training and knowledge exchange hub, enhance research excellence, and address the prevailing research and innovation gap within Widening countries.

The training program was structured into two key sessions:

- ChatMED Opportunities (April 25-26, 2025): This session integrated foundational principles of Generative AI and insights into establishing collaborative relationships with industry, including all the original themes of the Training Session 1 - Generative AI Basic, Hands-on Generative AI, Networking & Collaboration and Infrastructure & HPC.
- Advanced AI Applications in Healthcare (July 9, 2025): The themes on Domain-driven software design, User experience in medical applications, AI & Neurological Disorders and Medical imaging & AI were addressed in full.

The following Agenda presented in Figure 1 presents the official agenda for the ChatMED Opportunities event by including all the presentations in the plan (later partly postponed for 9th of July).



Figure 1. ChatMED Opportunities Agenda.

For each of the 36 confirmed participants we prepared a name tag specifically tailored for the event with an example given in Figure 2.



Figure 2. Name Tag example for the ChatMED Opportunities event.

The actual implementation of the planned topics within D2.4 is given in the following Table 1. Each of the links lead to the unique presentation uploaded at the project website.

Table 1. Training sessions descriptions.

Planned Topic (D2.4)	Actual Presentation (Presenter) (Link)	Type	Key Coverage / Highlights
Generative AI Basics	Generative AI Trends ( <i>Dimitar Trajanov</i> ) ( <a href="https://chatmed-project.eu/knowledge-repository/training-session-1-generativeai-trends/">https://chatmed-project.eu/knowledge-repository/training-session-1-generativeai-trends/</a> )	Theoretical presentation	Introduction to core concepts, LLM models, AI agents, communication protocols (MCP, A2A), and ethical/societal implications of Generative AI.
Hands-on Generative AI	Prompt Engineering for Healthcare Applications ( <i>Ana Todorovska</i> ) ( <a href="https://chatmed-project.eu/knowledge-repository/training-session-1-hands-on-generative-ai/">https://chatmed-project.eu/knowledge-repository/training-session-1-hands-on-generative-ai/</a> )	Practical session with examples	Explanation of prompt engineering techniques for healthcare, including manual and automatic methods, plus details about structuring it for better results.
Networking & Collaboration	ChatMED and the National Health Strategy up to 2030 ( <i>Monika Simjanoska Misheva</i> ) ( <a href="https://chatmed-project.eu/knowledge-repository/">https://chatmed-project.eu/knowledge-repository/</a> )	The entire sessions was dedicated to promoting and explaining various opportunities.	



	<a href="#">y/chatmed-and-the-national-healthstrategy-up-to-2030-how-genai-can-improve-the-health-system/) – part 1</a>		
Infrastructure & HPC	How GenAI can improve the health system? ( <i>Kostadin Mishev et al.</i> ) ( <a href="https://chatmed-project.eu/knowledge-repository/chatmed-and-the-national-healthstrategy-up-to-2030-how-genai-can-improve-the-health-system/">https://chatmed-project.eu/knowledge-repository/chatmed-and-the-national-healthstrategy-up-to-2030-how-genai-can-improve-the-health-system/</a> ) – part 2	Session showing practical examples and latest results from the researchers.	
Domain-driven software design	Domain Driven Design: Tailoring intelligent LLM Agents Based on [Medical] Domain Needs and Challenges ( <i>Riste Stojanov</i> ) ( <a href="https://chatmed-project.eu/knowledge-repository/training-session-2-domain-driven-design/">https://chatmed-project.eu/knowledge-repository/training-session-2-domain-driven-design/</a> )	Theoretical.	
Medical Imaging & AI	Medical Imaging & AI ( <i>Ivan Kitanovski</i> ) ( <a href="https://chatmed-project.eu/knowledge-repository/training-session-2-medical-imaging-ai/">https://chatmed-project.eu/knowledge-repository/training-session-2-medical-imaging-ai/</a> )	Practical, hands-on demo	Explanation of each medical techniques and usage along with a real demo that the participants could understand which parts the solution can contribute to help.
AI & Neurological Disorders	AI & Neurological Disorders ( <i>Katarina Trojchanec Dineva</i> ) ( <a href="https://chatmed-project.eu/knowledge-repository/training-session-2-ai-neurological-disorders/">https://chatmed-project.eu/knowledge-repository/training-session-2-ai-neurological-disorders/</a> )	Theoretical and survey-oriented	Overview of AI for assisting with all steps on diagnosing and treating Alzheimer's, Parkinson's, Stroke, and Neurological Cancers.
User experience in medical applications	Revolutionizing Multimodal Knowledge Retrieval ( <i>Jovana Dobрева</i> ) ( <a href="https://chatmed-project.eu/knowledge-repository/training-session-2-revolutionizing-multimodal-knowledge-retrieval/">https://chatmed-project.eu/knowledge-repository/training-session-2-revolutionizing-multimodal-knowledge-retrieval/</a> )	Presentation of techniques to better leverage multimodal information	GraphRAG, and general RAG techniques, for an optimal LLM solution.

The focus of the original ChatMED Opportunities event transformed into a strategic move to boost collaboration, trigger partnerships, and showcase industry engagement. With this decision, a special industry panel was introduced. The panel focused on opening up discussions on the possibilities for collaboration, spreading the consortium in novel projects and actual results delivered by the HPC equipment at the faculty. The panel comprised a diverse group of industry representatives, including from: Sorsix, RLDatix, Ekonet, Zan Mitrev Clinic, and Axians. Each representative was invited with a reason and this reason was clearly presented and is present in the slides dedicated to the topic Networking & Collaboration in Table 1.

The panel proved successful in initiating memorandums of collaboration which will be delivered in the first report scheduled for January 2027. As such, this event allowed a deeper exploration of existing industry needs and opportunities within the context of Generative AI, enriching the participant experience and setting the stage for future collaborations. Two of those collaborations are foreseen to be brought to light via mutual work on two project proposals and joint papers. Details will be presented in the upcoming progress report.

## Report on training sessions

The following section presents summaries/highlights of the presentations in the order as they were executed.

### **Training Session 1: Foundations of Generative AI**

#### **Overview:**

The initial session focused on introducing participants to the fundamental principles of Generative Artificial Intelligence (AI). The curriculum included an in-depth overview of contemporary generative models and their applications across various domains, with special emphasis on healthcare and academic-industry collaboration.

#### **Networking & Collaboration: Prof. Monika Simjanoska Misheva (PI)**

Prof. Misheva opened the training sessions targeting the topic of Networking & Collaboration. The presentation presented was tailored for the industry panel, aiming at showing the potential of the project to get involved with the other companies' solutions.



The idea with the presentation “*ChatMED and the National Health Strategy up to 2030*” was to position the project within the national health strategy, pointing out the weak points that can be significantly improved. The industry panel had representatives from:

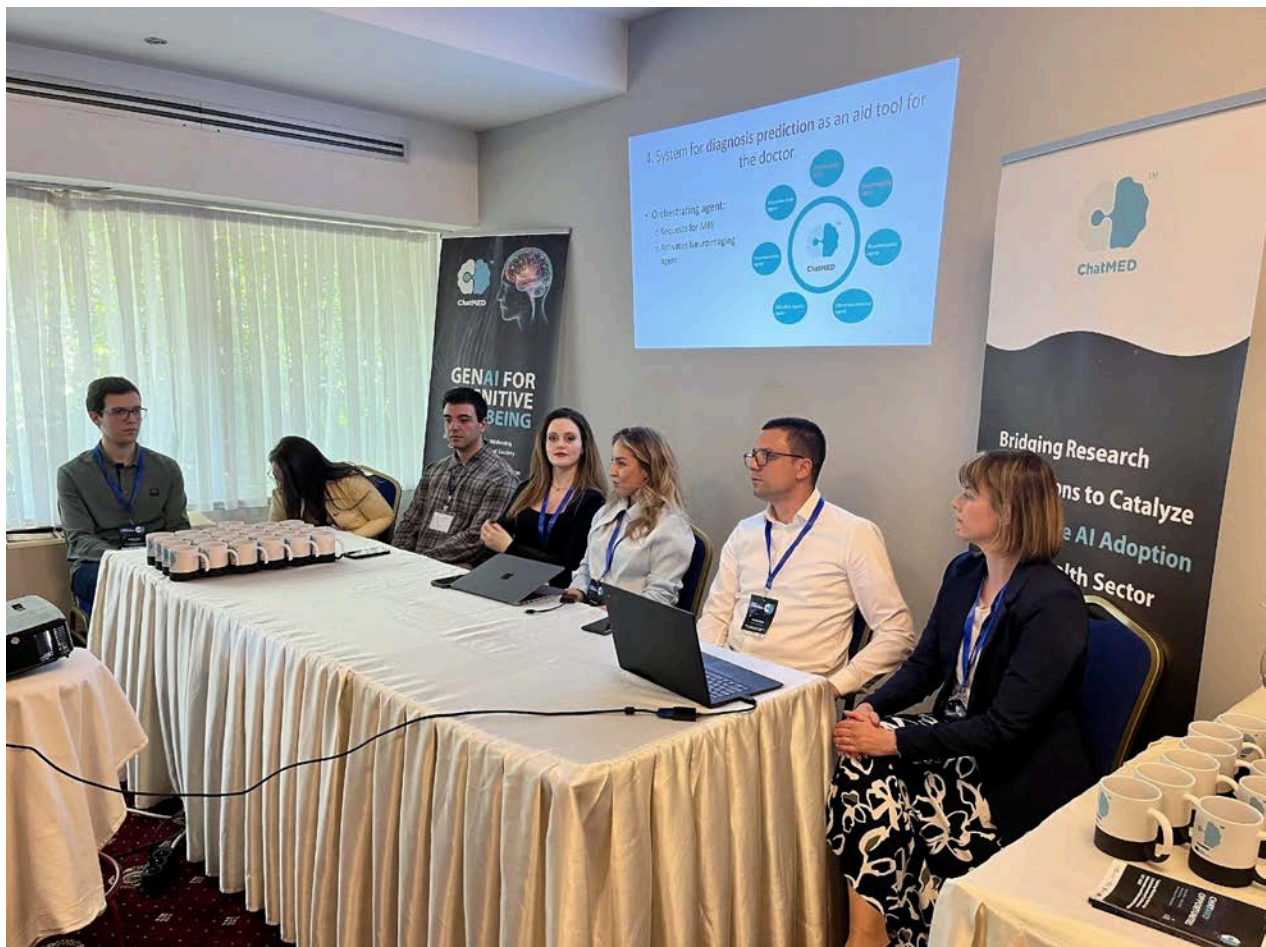
- **Sorsix** – the vendor for the national health system, encompassing the primary, secondary, and tertiary health (N. Macedonia is one of the rare countries to have such wide and interoperable health systems). Sorsix is also holding part of the health systems in Australia and New Zealand. It has been presented how ChatMED can fit in their Pinga system, embedding diagnosis prediction models as clinical decision support tools within national health records. The potential for using automatic transcription modules to streamline doctors' workflows and enhance smart retrieval capabilities for patient histories were also explored.
- **RLDatix** – RLDatix Catalix is a solution that concentrates on governance, risk, compliance, and workforce management. The ChatMED project has been proposed to integrate auto-logging narrative notes into structured records via automatic transcription modules, helping to reduce staff workload and burnout by automating low-risk interactions with voice bots. There is a potential integration for clinical risk surveillance systems to proactively flag high-risk or rare conditions.
- **EKONET** - With EKONET's experience alongside Kontron, a major embedded solutions provider in Macedonia, Slovenia, and Germany, ChatMED could explore innovative embedded devices with offline support for doctors, enabling "usage on the fly." The discussion involved embedding TTS, STT, and recommendation models directly into these devices, bringing AI-powered healthcare support directly to the field.
- **Axians** - Representing VINCI Energies with a strong presence across 35 countries, Axians can support ChatMED solutions through hospital-grade network reliability, end-to-end telecom solutions adaptable to hospital environments, secure virtual environments, HIPAA/GDPR-compliant hosting solutions, and reliable scalable data storage for EHRs, imaging, and analytics.

In summary, the presentation outlined five promising use cases that could emerge from ChatMED in the near future. These use cases have the potential to be integrated as pilot components within existing healthcare solutions, contingent upon the signing of a memorandum of cooperation and a non-disclosure agreement to establish mutual collaboration. The proposed use cases include:

- Automatic transcription of doctors' reports.
- Integration of voice bot for giving general patient-safe recommendations.
- Integration of system for diagnosis prediction as an aid tool for the doctors.
- Integration of system for making longitudinal summary reports for a patient.
- Integration of system for smart retrieval across patients' histories and literature.

### **Infrastructure & HPC: Prof. Kostadin Mishev et al. (Scientific Committee)**

Prof. Mishev continued the presentation by demonstrating how the project's infrastructure has already enabled remarkable results within the first eight months of ChatMED's implementation. Each team shared initial insights from specific case studies, illustrating the tangible outcomes of early-stage development.



The presentation began with an overview of the infrastructure capacity, highlighting a robust configuration comprising 928 GB of total VRAM, powered by state-of-the-art GPU resources: eight A100s, one H100, and a range of Quadro RTX units. This setup provides the computational foundation necessary for training and deploying large-scale AI models.



Following this, the teams showcased their progress in model development and evaluation. Recognizing the importance of cultural and linguistic relevance, a central objective was the development of VezilkaLLM, which is the first-ever large language model tailored to the

Macedonian language. The model was evaluated across diverse benchmarks, validating its performance and usability in local medical contexts.



Subsequently, the presentation introduced a high-performance speech-to-text (STT) system in Macedonian, integrated within a VoiceBot platform developed by the company iReason. Offered as an in-kind contribution to FCSE, this platform supports rapid prototyping and deployment of conversational agents. It was demonstrated how this infrastructure can support a range of use cases, from automating the transcription of doctors' reports to enabling general patient-safe medical recommendation VoiceBots.

One of the most impactful segments featured a diagnosis prediction use case, where a clinician-facing assistant leverages a Retrieval-Augmented Generation (RAG) pipeline to provide evidence-backed suggestions. This case study was developed in collaboration with clinical experts from Zan Mitrev Clinic, which is now an official partner in forthcoming joint projects.

The next section addressed the role of multimodal large language models (MLLMs) in neuroimaging. By incorporating expert-annotated brain MRI data, it was shown how fine-tuning significantly improves model accuracy and clinical relevance.

To facilitate the involvement of domain experts in result evaluation, a demonstration of the Human Evaluation Platform (currently under development) was presented. The demo showcased the interface and key functionalities that will serve as the foundation for selecting suitable LLMs for further fine-tuning. The final demonstration illustrated the capacity of LLMs to process and interpret EMG signals for the prediction of neurological conditions such as ALS and

myopathy. This was achieved through a comparative analysis of both zero-shot and fine-tuned LLM-based classifiers, emphasizing the diagnostic potential of these models.

In conclusion, the session emphasized a key paradigm shift: rather than relying on a single, all-knowing LLM, the future lies in multi-agent ecosystems, where specialized LLMs, each trained on a narrow domain or modality, collaborate to solve complex, real-world problems in healthcare.

### **Generative AI Basics: Prof. Dimitar Trajanov (Scientific Committee)**

Prof. Dimitar Trajanov has adapted the topic to “Generative AI Trends”, however, introducing the foundational concepts of Generative AI, highlighting its significance in various fields and its potential to revolutionize healthcare. Focused on distinguishing between different classes within data, modeling the conditional probability of the output given the input, creating new content by understanding and replicating patterns in training data and highlighted several examples.

- Best LLM models: The segment analyzed the best models, listing generalist models and vision models with their rankings. It touched on their cost, and performance tradeoffs.
- The second wave of GenAI: Prof. Trajanov highlighted the rise of AI agents, focusing on their ability to automate complex, multi-step workflows and make informed decisions. Highlighted both Autonomous Agents, and A2A protocols.
- AI agent ecosystem: The discussion on AI agents extended to explore Agent Communication Protocols, including Model Context Protocol (MCP) and Agent2Agent (A2A), underscoring the importance of standardized communication methods for multi-agent collaboration. Covered what it is, how it works, its core components.
- Impact of GenAI to society: The societal implications of Generative AI, encompassing Workforce Evolution and the Information Ecosystem & Misinformation, were examined, outlining its key impacts on labor productivity, personalizing customer experience, accelerating R&D, and new business models.



### Hands-on Generative AI: Ass. Ana Todorovska (PhD Student, Researcher)

Ass. Ana Todorovska adapted the topic to “Prompt Engineering for Healthcare Applications”. Instead of retraining LLMs from, prompt engineering allows practitioners to shape AI behavior through carefully crafted inputs. This process, known as prompt learning, enables even general-purpose models to carry out highly specific medical tasks by refining how questions are asked. Prompt learning follows an iterative cycle: select a capable pre-trained model, design task-specific prompts, evaluate the outputs, and refine the prompts accordingly. Through this process, developers and clinicians can align model outputs with real-world clinical needs, without accessing protected datasets or altering the model's core parameters.



Two primary methods for designing prompts are presented:

- Manual prompting: crafted by human experts, often based on clinical insight.
- Automatic prompting: generated and optimized by algorithms, allowing for exploration at scale.

Both manual and automatic prompting encompass techniques such as:

- Zero-shot prompting: issuing direct instructions without examples.
- Few-shot prompting: including example pairs to guide the model's response style.
- Prompt mining: extracting patterns from medical literature to generate useful prompts.
- Prompt paraphrasing and scoring: using other AIs to improve and evaluate prompt quality.
- Continuous prompting: using embeddings instead of plain text for finer control.

Given the high-stakes nature of healthcare, random experimentation with prompts is not viable. Instead, structured frameworks like Prompt-Eng have emerged. Inspired by clinical trial methodology, Prompt-Eng offers a systematic way to test and improve prompt strategies. It emphasizes goal definition, metric selection, real-world evaluation, and iterative refinement. In practice, using a set of diverse prompts rather than a single formulation tends to yield more robust results. This redundancy allows the model to consider multiple perspectives, reduces bias, enhances generalizability, and increases clinician confidence through comparative review of outputs.

An advanced application of prompt engineering is RAG which enhances model performance by grounding answers in real-time information retrieved from trusted external sources. The process begins with a user query, which is expanded and clarified through prompt engineering techniques. Relevant documents are retrieved, and the final prompt combines the query, the retrieved context, and instructions such as citing sources or prioritizing certain evidence types.

During the hands-on session, participants observed how these principles were applied using the Gemma-3 model. Real clinical examples demonstrated that while current outputs are promising, further refinement is required to meet production-level safety and accuracy standards. Doctors confirmed that prompt quality directly influenced clinical usability, validating the importance of continued iteration.

The frontier of multi-modal prompting is also expanding, integrating text, images, and sensor data. Reinforcement learning techniques are enabling prompts to be optimized based on clinician feedback, ensuring continuous improvement. Meanwhile, embedded or trainable prompts, where prompt structures are treated as learnable parameters inside the model, may revolutionize customization.

## **Training Session 2: AI for Neurological Applications**

### **Overview:**

The second session centered on the development and customization of AI-driven solutions to address neurological challenges. The session promoted a user-focused design neurology, emphasizing applications tailored for clinicians and patients in the neurological domain.

### **Domain-Driven Software Design: Prof. Riste Stojanov (Technical Committee)**

Prof. Stojanov has adapted the topic to “Domain Driven Design: Tailoring intelligent LLM Agents Based on [Medical] Domain Needs and Challenges”. The presentation focused on applying the principles of Domain-Driven Design (DDD) to the development of intelligent LLM agents tailored for complex medical tasks. As generative models like ChatGPT, Claude, and others become increasingly integrated into clinical workflows, ensuring domain specificity, safety, and contextual accuracy is paramount. DDD provides a structured methodology to achieve this.



A core component of DDD is the use of Bounded Contexts - clearly defined modules that encapsulate a part of the domain model. Each context is associated with a specific subdomain and vocabulary, allowing LLMs to be segmented and specialized for discrete clinical functions. This modularity is especially useful for multi-agent systems, where each agent can represent a bounded context and communicate through predefined interfaces.

The presentation emphasized the importance of Ubiquitous Language, a shared vocabulary co-developed by developers and clinicians. This common language ensures that prompts, outputs, validation rules, and tool interfaces are consistently interpretable by both human and machine. By using consistent terminology, LLMs can better align their outputs with domain expectations and reduce semantic ambiguity.

To address the challenge of hallucinations, the presentation proposed the use of Value Objects and Structured Outputs. These mechanisms constrain the model's output to predefined schemas (e.g., HemoglobinLevel, DiagnosisCode), enabling more reliable interpretation and downstream validation.

For scenarios where LLMs encounter limitations in internal knowledge, the session introduced RAG to enable LLMs to fetch up-to-date, task-specific knowledge from external sources (e.g., PubMed, clinical repositories) and incorporate it into their reasoning processes. This hybridization of internal and external knowledge significantly boosts clinical applicability while enhancing trust and traceability.

The session also touched on reasoning decomposition for complex medical tasks. Instead of relying on a monolithic prompt, the system can plan multiple intermediate reasoning steps, execute tools iteratively, and then compile a final response. This planning mechanism helps agents maintain an internal state and execute context-sensitive logic, such as calculating dosage adjustments or interpreting lab panels across time.

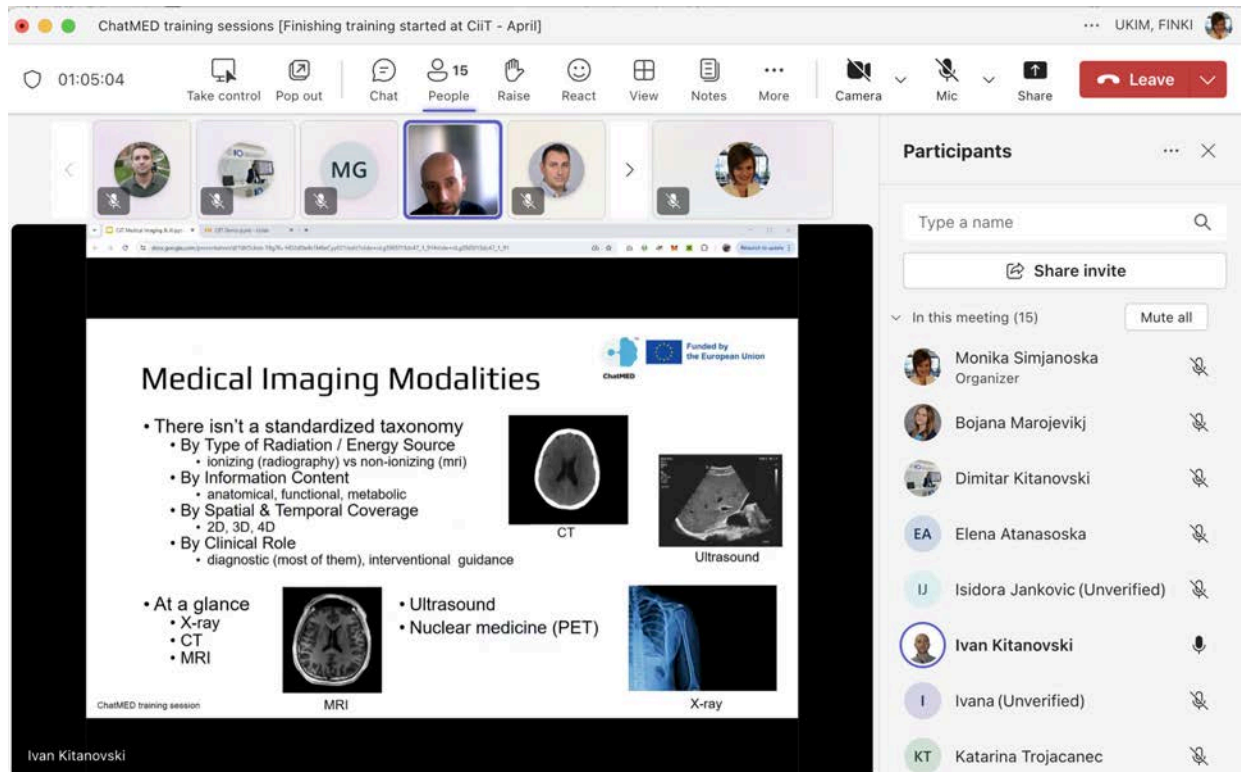
In practical terms, the Model Context Protocol (MCP) was introduced as a standard to interface LLM agents with external tools and databases. MCP allows for interoperability between agents and structured data sources (e.g., local files, SQLite, web APIs), forming the communication backbone of the multi-agent architecture presented in the ChatMED platform.

The session concluded by summarizing how DDD's constructs map to LLM agent systems: bounded contexts define modular agents, ubiquitous language harmonizes inputs and outputs, aggregates ensure planning integrity, and repositories act as anti-corruption layers connecting LLMs to clinical systems.

### **Medical Imaging & AI: Prof. Ivan Kitanovski (Neuroimaging Researcher)**

The training session on Medical Imaging & AI, led by Ivan Kitanovski, provided a practical and insightful exploration of how AI is transforming radiology and diagnostic imaging. Kitanovski began by reviewing the diverse modalities used in clinical imaging: X-rays, CT (Computed Tomography), MRI (Magnetic Resonance Imaging), ultrasound, and PET scans. These were

grouped according to their energy sources (ionizing versus non-ionizing), the type of information they capture (anatomical, functional, or metabolic), and their spatial or temporal coverage (2D, 3D, 4D). Each modality plays a distinct role in clinical practice, from diagnostics to interventional guidance. One of the session’s highlights was the detailed comparison between MRI and CT. MRI, which uses magnetic fields and radio waves, provides exceptional soft-tissue contrast - ideal for imaging the brain, spinal cord, and joints. CT, on the other hand, relies on ionizing X-rays and is the preferred choice in trauma settings due to its speed and utility in detecting fractures, hemorrhages, and pulmonary embolisms. Participants were reminded of the importance of modality selection based on clinical context, urgency, and patient risk factors.



The screenshot shows a Zoom meeting interface. At the top, the title bar reads "ChatMED training sessions [Finishing training started at CiiT - April]". The meeting controls include a timer at 01:05:04, icons for Take control, Pop out, Chat, People (15), Raise, React, View, Notes, More, Camera, Mic, Share, and a Leave button. Below the controls is a gallery view of participants, with one participant highlighted. The main content area displays a slide titled "Medical Imaging Modalities" with the following text:

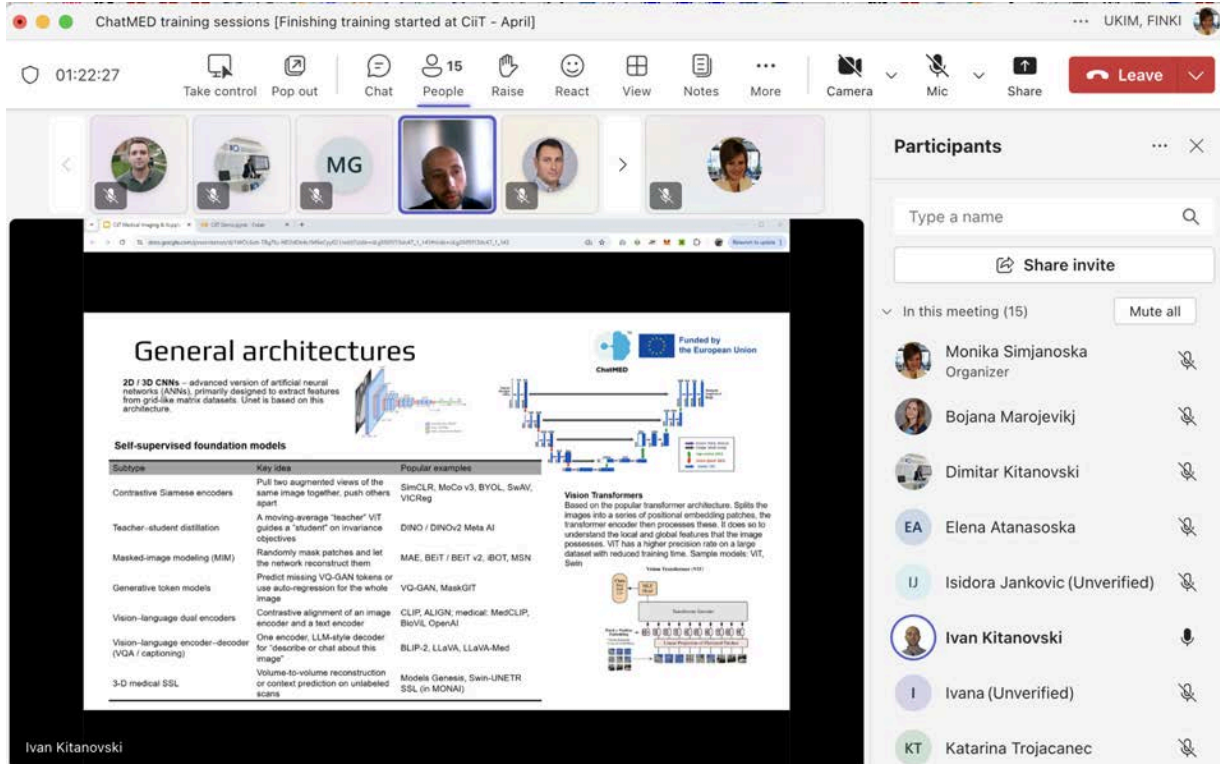
**Medical Imaging Modalities**

- There isn't a standardized taxonomy
  - By Type of Radiation / Energy Source
    - ionizing (radiography) vs non-ionizing (mri)
  - By Information Content
    - anatomical, functional, metabolic
  - By Spatial & Temporal Coverage
    - 2D, 3D, 4D
  - By Clinical Role
    - diagnostic (most of them), interventional guidance
- At a glance
  - X-ray
  - CT
  - MRI
- Ultrasound
- Nuclear medicine (PET)

The slide also features four images: a CT scan of a brain, an Ultrasound image, an MRI scan of a brain, and an X-ray of a shoulder. The bottom left of the slide is labeled "ChatMED training session" and "Ivan Kitanovski".

On the right side of the meeting, the "Participants" list shows 15 people in the meeting. The list includes:
 

- Monika Simjanoska (Organizer)
- Bojana Marojevikj
- Dimitar Kitanovski
- EA Elena Atanasoska
- IJ Isidora Jankovic (Unverified)
- Ivan Kitanovski
- I Ivana (Unverified)
- KT Katarina Trojancanec



**General architectures**

2D / 3D CNNs – advanced version of artificial neural networks (ANNs), primarily designed to extract features from grid-like matrix datasets. Unet is based on this architecture.

**Self-supervised foundation models**

Subtype	Key idea	Popular examples
Contrastive Siamese encoders	Pull two augmented views of the same image together, push others apart	SimCLR, MoCo v3, BYOL, SwAV, VICReg
Teacher-student distillation	A moving-average “teacher” ViT guides a “student” on invariance objectives	DINO / DINOv2 Meta AI
Masked-image modeling (MIM)	Randomly mask patches and let the network reconstruct them	MAE, BEiT / BEiT v2, iBOT, MSN
Generative token models	Predict missing VQ-GAN tokens or use autoregression for the whole image	VQ-GAN, MaskGIT
Vision-language dual encoders	Contrastive alignment of an image encoder and a text encoder	CLIP, ALIGN, medical: MedCLIP, BioViL, OpenAI
Vision-language encoder-decoder (VQA / captioning)	One encoder, LLM-style decoder for “describe or chat about this image”	BLIP-2, LLaVA, LLaVA-Med
3-D medical SSL	Volume-to-volume reconstruction or content prediction on unlabeled scans	Models: Genesis, Swin-UNETR SSL (in MONAI)

**Vision Transformers**  
Based on the popular transformer architecture. Splits the images into a series of positional embedding patches, the transformer encoder then processes these. It does so to understand the local and global features that the image possesses. ViT has a higher precision rate on a large dataset with reduced training time. Samples: models, ViT, Swin

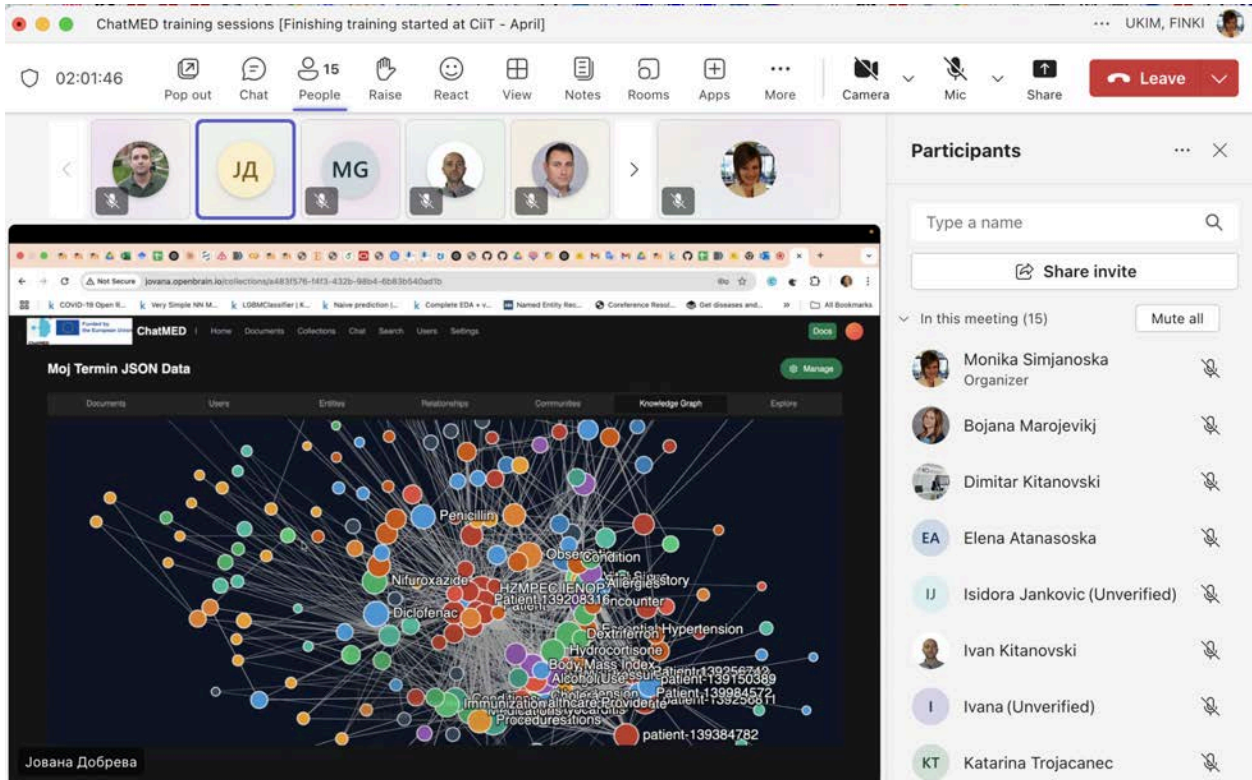
Ivan Kitanovski

The session moved from physics into data formats. Clinical imaging data is primarily stored in DICOM (Digital Imaging and Communications in Medicine) format, which bundles pixel-level image data with critical metadata like patient identifiers, study protocols, and imaging parameters. For research purposes, NIfTI files are common, particularly in neuroscience, offering streamlined storage of multi-dimensional imaging volumes. Other formats like TIFF, PNG, and JPEG are often used for report illustrations or training visual models, while cloud-native formats such as HDF5 and OME-NGFF are emerging for large-scale AI training. Kitanovski introduced the audience to the core tasks AI supports in medical imaging: classification, segmentation, detection, registration, and visual-language report generation. Real-world examples, such as automated tumor detection and multi-modal image alignment, were shown to highlight how these tasks reduce human burden while preserving diagnostic accuracy.

The deep learning architectures behind these applications were then unpacked. 2D and 3D convolutional neural networks (CNNs) form the foundation of most segmentation models, including the widely adopted UNet. More recent advancements involve Vision Transformers (ViTs), which apply attention mechanisms to capture long-range dependencies in image data. Self-supervised learning approaches, such as masked image modeling and contrastive pretraining, were presented as critical techniques for learning from unlabeled medical datasets. The integration of vision-language models (like CLIP and LLaVA-Med) was also explored, allowing AI systems to generate coherent radiology reports directly from image inputs.

In the final segment, Kitanovski showcased a hands-on demonstration (the demo script is available in the presentation) using a UNet model to perform brain tumor segmentation in real-time. The demo walked through the end-to-end pipeline: loading a medical scan,





The motivation for this work stems from a critical issue in healthcare AI: LLMs often hallucinate or misinterpret nuanced medical relationships when relying solely on unstructured text chunks. To solve this, Jovana proposed a new paradigm of knowledge retrieval that goes beyond vector similarity search. Her architecture, GraphRAG, incorporates document ingestion, entity extraction, graph-based storage, and query processing, enabling models to traverse structured relationships instead of relying only on lexical overlaps.

The session began with a deep dive into the four-layer GraphRAG pipeline. First is the Document Ingestion layer, which applies recursive chunking and OpenAI embeddings to process raw documents. The second step is Knowledge Graph Construction, where extracted medical entities are linked through relational inference and embedded into a graph database. These graphs provide semantic scaffolding for the model to understand complex clinical associations, such as gene-disease links or drug interactions. The third layer, Storage, maintains three parallel systems: a vector database for quick similarity search, a graph database for traversable context, and a document store for traceability. Finally, the Query Processing layer involves multiple LLMs, such as DeepSeek, that interact with graph traversal logic, semantic search, and context assembly engines. These components work together to generate responses grounded not just in relevant documents but in inferred, medically validated relationships. Dobрева used a compelling case study to illustrate the value of this system. When asked about the potential impact of a *PSEN1* genetic variant (c.1234A>G) on Alzheimer's pathology, a standard RAG model returned loosely relevant snippets. In contrast, GraphRAG successfully mapped the query to known entity relationships within the knowledge graph, linking *PSEN1*, amyloid-beta processing, and early-onset Alzheimer's, and produced a structured

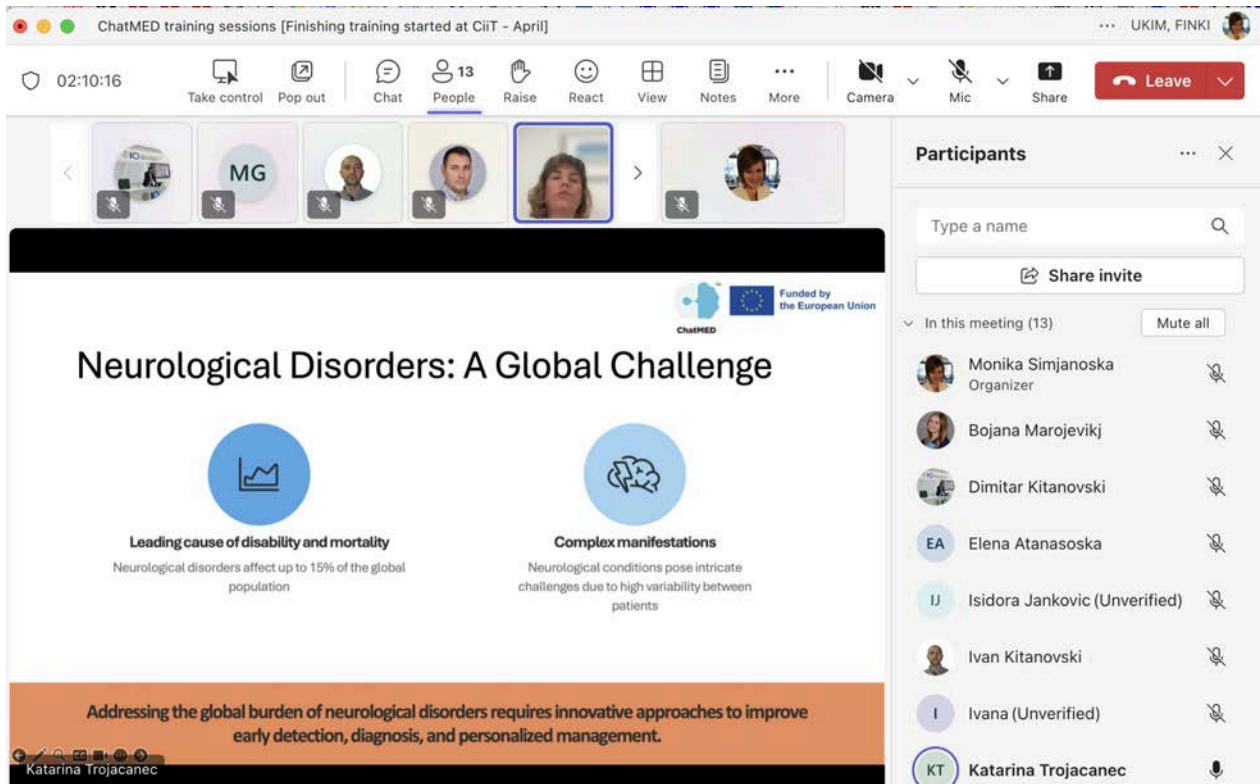
clinical hypothesis with citation trails. This example demonstrated not only improved accuracy but also a level of explainability essential for clinician trust.

One of the most visually informative slides compared traditional RAG against GraphRAG side-by-side. While standard RAG offers surface-level retrieval based on similarity scores, GraphRAG leverages relationship paths to deliver reasoning-informed answers rather than isolated facts.

Jovana closed by emphasizing the importance of integrating GraphRAG into platforms like ChatMED, where clinical reasoning, multimodal inputs, and human-in-the-loop validation are essential. Her work demonstrates that structured knowledge representation is not only possible but necessary to scale AI in medicine responsibly and effectively.

### AI & Neurological Disorders: Prof. Katarina Trojchanec Dineva (Neuroimaging Researcher)

Prof. Dineva explored in depth the transformative potential of AI in neurology. The session addressed the motivations behind using AI in this domain, practical applications, foundational technologies, and the challenges that remain for clinical adoption.



The screenshot shows a Zoom meeting interface. The main content is a slide titled "Neurological Disorders: A Global Challenge" with the following text:

**Neurological Disorders: A Global Challenge**

**Leading cause of disability and mortality**  
Neurological disorders affect up to 15% of the global population

**Complex manifestations**  
Neurological conditions pose intricate challenges due to high variability between patients

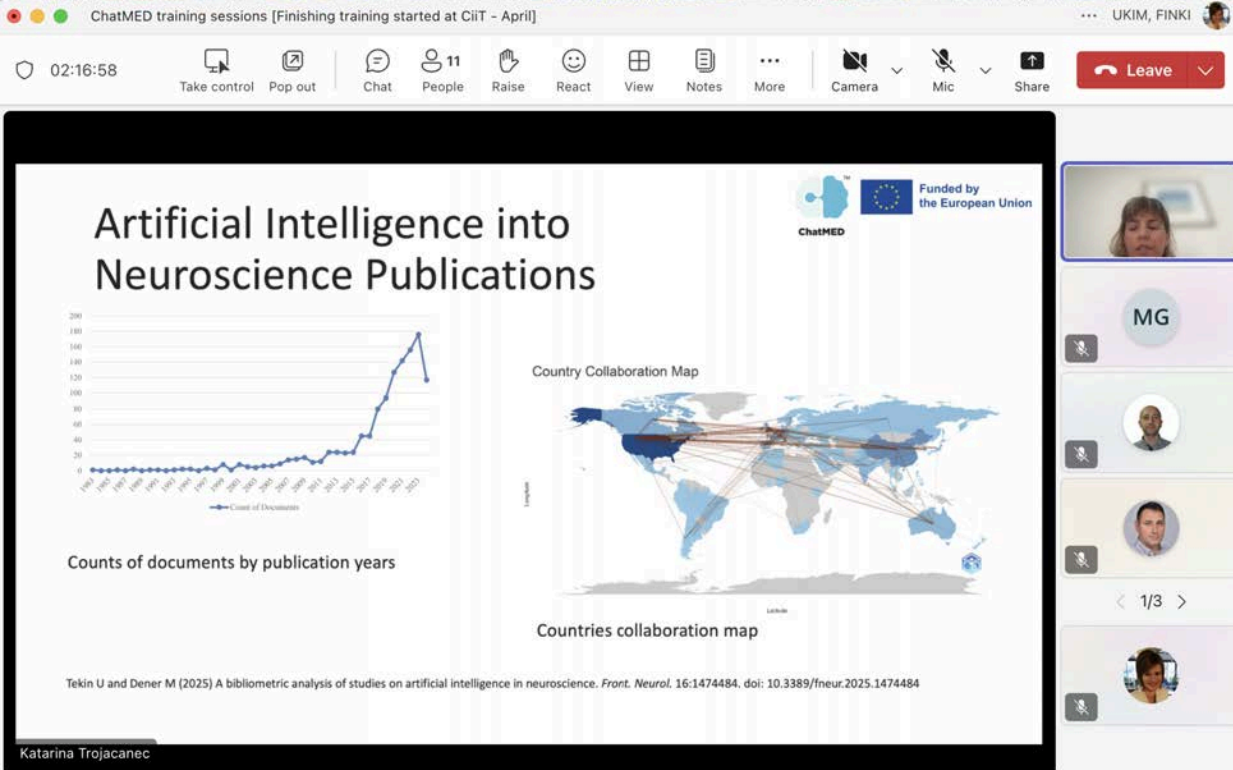
Addressing the global burden of neurological disorders requires innovative approaches to improve early detection, diagnosis, and personalized management.

The slide also features the ChatMED logo and the text "Funded by the European Union".

The participants list on the right includes:

- Monika Simjanoska (Organizer)
- Bojana Marojevikj
- Dimitar Kitanovski
- EA Elena Atanasoska
- IJ Isidora Jankovic (Unverified)
- Ivan Kitanovski
- I Ivana (Unverified)
- KT Katarina Trojchanec





**Artificial Intelligence into Neuroscience Publications**

Counts of documents by publication years

Country Collaboration Map

Countries collaboration map

Tekin U and Dener M (2025) A bibliometric analysis of studies on artificial intelligence in neuroscience. *Front. Neurol.* 16:1474484. doi: 10.3389/fneur.2025.1474484

Katarina Trojancanec

A pivotal advancement in the field has been the integration of foundation models. These models facilitate multimodal fusion, where text, images, and biological signals are combined into unified representations of patient profiles. This shift not only enhances diagnostic clarity but also enables holistic, patient-centric care. Real-world applications are already demonstrating promise. AI is being developed to improve detection of multiple sclerosis lesions, predict Alzheimer's progression, and enable faster identification of stroke in both pre-hospital and in-hospital settings. However, successful implementation in practice demands more than technological sophistication.

Several barriers remain. Data quality and representativeness are ongoing concerns, biased or non-diverse training data can lead to inequitable outcomes. The lack of explainability in black-box AI systems erodes trust among clinicians. Integration into clinical workflows must be seamless to prevent disruption. Finally, compliance with ethical and regulatory standards, including data privacy, transparency, and fairness is non-negotiable. To advance the adoption of AI in neurology, several actions are recommended: implementing standardized data sharing protocols; developing explainable AI (XAI) models; embedding ethical frameworks around consent and equity; fostering interdisciplinary collaboration; and ensuring continuous validation in real-world settings.

In conclusion, AI has the potential to act not as a replacement, but as a collaborative partner in neurological care. When deployed with clinical insight and ethical rigor, AI can empower neurologists to make more informed, timely, and personalized decisions, shifting neurology from reactive disease management to anticipatory, integrated, and patient-centered care. The future of neurology is not only intelligent but collaborative.



# Report on STSE

## STSE Report for Topic 1

Neurological Case Studies: Real-world Applications, Challenges, and Solutions in AI-Powered Neurology

**Name of Staff Members:**

Assistant Dimitar Kitanovski, Young Researcher (PhD Student)

Assistant Ana Todorovska, Young Researcher (PhD Student)

**Host Institution:**

Jožef Stefan Institute (JSI), Ljubljana, Slovenia

**Exchange Period:**

12 May – 31 May 2025 (14 working days)

**Overview**

Under the ChatMED Work Package on capacity building and in accordance with Deliverable D2.4 (“Detailed Training Curriculum and Talent Mobility Programs”), the Short-Term Staff Exchange (STSE) on Topic 1 was designed to investigate the design and implementation of AI-based systems for guiding users through authentic neurological case studies. The activity, hosted at JSI, directly addressed the task of developing a multi-agent orchestration system that supports clinical reasoning in neurology by integrating textbook knowledge with real-world medical scenarios. Both Ana Todorovska and Dimitar Kitanovski participated actively in orchestrator workflow planning, dataset preparation, and the specification of modular agent functionalities, thereby laying the groundwork for the system’s initial prototype.

**Objective Alignment with Topic 1**

The STSE objectives corresponded precisely to those defined in D2.4 for Topic 1: to develop a large-language-model–based orchestrator agent capable of guiding users through neurological case studies by retrieving, analysing, and recommending subsequent steps based on user input, leveraging both canonical procedures and empirical clinical data. During the exchange, the team achieved the following milestones: structuring neurological workflows derived from domain-expert–validated textbook material; creating a case-based dataset complete with metadata for seamless LLM integration; designing the logic and modular architecture of the multi-agent system; and implementing a preliminary web interface to facilitate future testing and validation. These accomplishments fulfil the planned early-phase deliverables and position the orchestrator for completion by the deadline of the first progress report.

**Description of Work Performed**

The STSE unfolded over 14 working days in a structured, collaborative format, with guidance from the JSI FMN neuro-informatics group and senior FCSE researchers (Prof. Kostadin Mishev, Prof. Katarina Trojancanec Dineva, and Prof. Monika Simjanoska Misheva).

### **Days 1–2: Project Planning and Resource Setup**

The researchers co-designed the system architecture, defining the orchestrator’s core functions user-input interpretation, relevant case retrieval, and diagnostic or therapeutic recommendation. Ana compiled neurology handbooks and journal case studies with FMN support, while Dimitar configured the development environment and extraction tools for structured content ingestion.

### **Days 3–5: Data Collection and Preprocessing**

Ana curated and cleaned a dataset of real-world neurological cases spanning multiple sclerosis, stroke, tumour pathology, epilepsy, and neurodegenerative disorders encoding each record in JSON with metadata fields for diagnosis, symptomatology, and treatment pathways. Concurrently, Dimitar extracted procedural protocols from neurology textbooks (PDF), organizing standard examination and intervention steps into a structured knowledge base for the orchestrator’s reference.

### **Days 6–8: Orchestrator Workflow Design**

Both researchers mapped the orchestrator’s decision logic: Ana formulated a stepwise reasoning framework linking user-provided clinical data to actionable diagnostic suggestions, grounded in accepted clinical workflows; Dimitar integrated structured textbook protocols into the backend logic, ensuring consistency between expert-defined procedures and automated inference.

### **Days 9–11: Dataset Preparation and Scenario Modeling**

Ana developed representative interaction scenarios such as symptom entry or test result reporting—to inform subsequent fine-tuning and evaluation phases. Simultaneously, Dimitar engineered the data pipelines enabling the orchestrator to parse user input, retrieve pertinent knowledge, and generate contextually appropriate responses.

### **Days 12–14: Multi-Agent Design and Testing**

The final phase focused on modular multi-agent architecture: Ana and Dimitar defined distinct roles for retrieval, diagnosis, and treatment-planning agents, and Dimitar proposed a coordination framework for inter-agent communication. Ana conducted preliminary validation by simulating clinical inputs to assess the coherence and clinical relevance of orchestrator outputs.

### **Outcome and Impact**

The STSE yielded multiple key deliverables: a richly annotated, case-based dataset suitable for LLM fine-tuning and evaluation; a structured repository of procedural clinical sequences extracted from textbooks; a multistep decision-logic design for the orchestrator (illustrated in Figure 4); a high-level multi-agent system architecture with clearly delineated agent responsibilities; and a prototype web interface established for future expert validation and usability testing. Collectively, these outputs provide a robust foundation for advancing LLM fine-tuning, full system integration, and clinical validation in subsequent project phases.

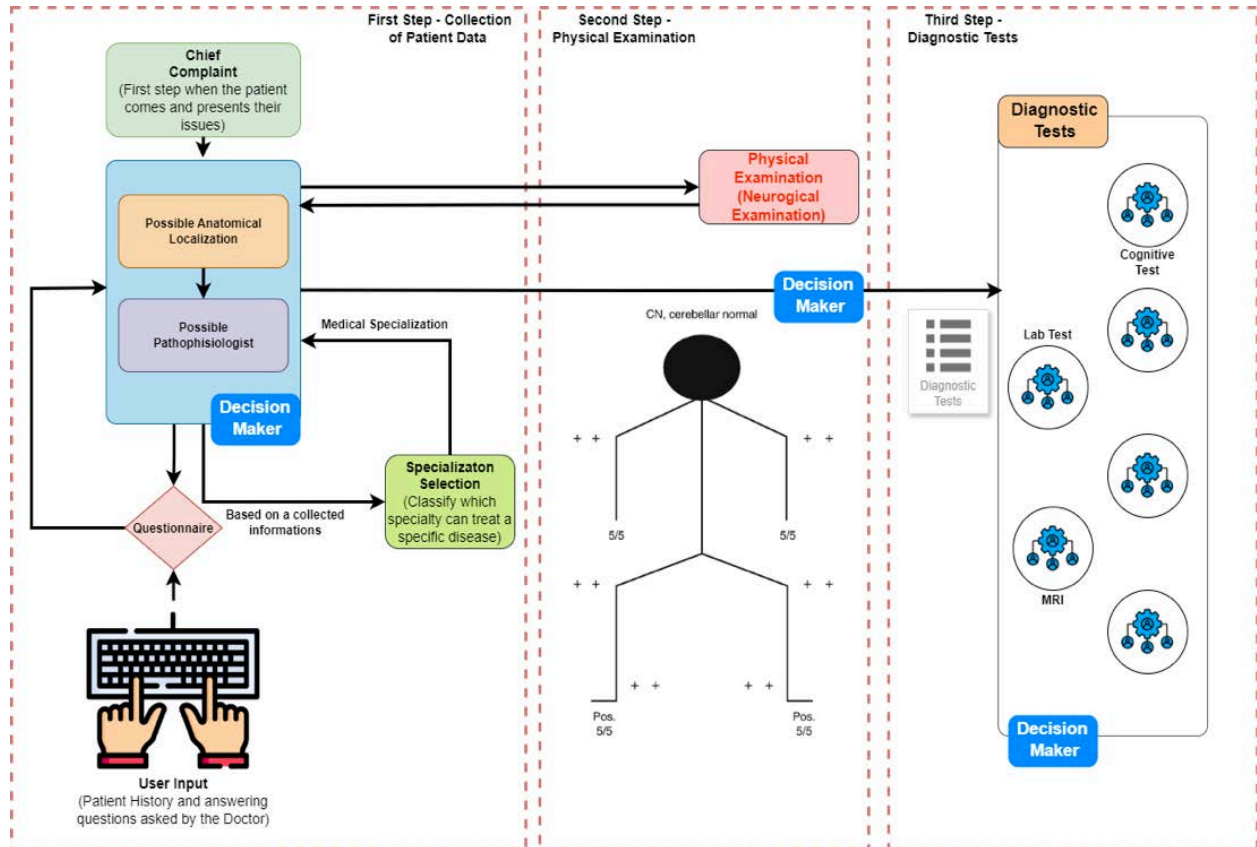


Figure 4. Workflow design for the neurology-driven orchestrator.

## Conclusion

The short-term exchange successfully fulfilled the Year 1 objectives of Topic 1 by delivering comprehensive design, data-curation, and integration frameworks for an AI-powered neurological case-study assistant. The collaborative efforts of Ana Todorovska and Dimitar Kitanovski, supported by domain experts at JSI and FCSE, have set the stage for the next stages of ChatMED development. The complete codebase for the Neuro-Orchestrator is publicly accessible at <https://github.com/ChatMED/llm-neurochestration/tree/main>.

## STSE Report for Topic 2

Advanced Medical Imaging Techniques: Understanding State-of-the-Art Imaging Modalities and Their Data Interpretations

### Name of Staff Member:

Prof. Katarina Trojchanec Dineva, Senior Researcher

### Host Institution:

Jožef Stefan Institute (JSI), Ljubljana, Slovenia

**Exchange Period:**

12 May – 31 May 2025 (14 working days)

**Overview:** During the Short-Term Staff Exchange at the Jožef Stefan Institute, Prof. Trojchanec Dineva concentrated on integrating advanced neuroimaging analysis capabilities into the ChatMED multi-agent system. The principal aim was to enable multimodal clinical reasoning by leveraging vision-enabled large language models (MLLMs) capable of jointly interpreting medical images and associated clinical text. To this end, state-of-the-art models including OpenAI GPT variants, MedGemma, LLaVA-Med, and others were evaluated on clinically relevant datasets covering brain cancer, multiple sclerosis, and stroke. The core deliverable was a neurological image analysis agent able to process two-dimensional slices, as well as three- and four-dimensional imaging volumes from MRI, CT, and PET modalities, and to communicate its findings to the orchestrator and peer agents for collaborative decision-making.

**Objective Alignment with Topic 2:** This work directly addressed Deliverable D2.4's mandate to integrate advanced imaging techniques into an AI-driven decision-support framework. In particular, the objective to design an image analysis agent capable of interpreting MRI, CT, and PET data and participating in a multi-agent orchestration aligns precisely with the STSE's activities. By benchmarking MLLMs on curated neuroimaging datasets and integrating the resulting agent within the existing orchestration architecture, the exchange fulfilled the implementation and system-level integration targets defined for Topic 2.

**Description of Work Performed**

**Days 1–2: Preparatory Planning:** The initial phase involved defining functional requirements for the imaging agent and selecting representative datasets. Conditions of interest brain cancer, multiple sclerosis, and stroke were chosen to span a range of classification complexities. Publicly available neuroimaging repositories with complete annotations were identified alongside expert-provided case studies and neurology handbooks, ensuring both quality and clinical relevance for subsequent fine-tuning and validation.

**Days 3–7: Agent Architecture and Specification:** Prof. Dineva designed the image analysis agent's architecture, specifying its core responsibilities: ingestion of 2D, 3D, and 4D MRI, CT, and PET data; preprocessing pipelines (including resampling, skull-stripping, and modality harmonization); detection of neuropathological features (tumors, lesions, infarcts, atrophy); and transformation of findings into structured JSON outputs. Each output schema was defined to include diagnostic labels, ICD-10 codes, anatomical localization, rationale, image modality and plane, recommended next steps, and confidence scores. Input/output protocols were formalized to enable seamless integration with the orchestrator and peer agents.

**Days 8–11: System Integration and Collaborative Workflow:** Attention then shifted to embedding the imaging agent within the multi-agent orchestration framework. The orchestrator was configured to invoke image analysis workflows at appropriate junctures in the diagnostic process. Mechanisms for confidence-based routing were established, allowing the system to escalate ambiguous cases for human-in-the-loop review. Inter-agent message schemas were

tested to ensure that imaging insights, clinical context, and other diagnostics converged into coherent recommendations.

**Days 12–14: Benchmarking and Validation Planning:** A suite of vision-enabled LLMs OpenAI GPT, MedGemma, Gemini, Amazon Nova, and LLaVA-Med 1.5 was selected for comparative evaluation. Preliminary benchmarks were defined using publicly accessible MRI and CT datasets alongside clinician-curated cases. Metrics for medical context understanding, diagnostic accuracy, and reasoning coherence were established to guide future fine-tuning and performance assessments. The final day was devoted to documenting challenges such as DICOM versus NIfTI interoperability and volumetric data handling and drafting a roadmap for addressing them in subsequent project phases.

**Outcome and Impact:** By the end of the exchange, the team had completed the architectural design of a vision-enabled image analysis agent and defined its integration points within the multi-agent system. The agent’s schema and preprocessing pipelines are now in place, and the orchestration framework supports confidence-driven routing and multimodal data fusion. This infrastructure provides a flexible platform for rigorous benchmarking of emerging MLLMs in neurology and sets the stage for fine-tuning, clinical validation, and expansion to additional imaging modalities and neurological conditions.

**Conclusion:** The STSE successfully established the foundations for multimodal neuroimaging analysis within ChatMED’s AI-orchestrated decision-support system. Prof. Trojchanec Dineva’s work on dataset curation, agent specification, and system integration underscores the promise of combining structured agents with vision-enabled LLMs to enhance clinical reasoning in neurology. Future efforts will focus on volumetric data processing, interoperability enhancements, and extending diagnostic coverage to a broader spectrum of neurological disorders.

**Outcome:** Manuscript in preparation, entitled “Evaluation of Vision-Enabled Large Language Models for Clinical Reasoning in Neurological Disorders.”

## STSE Report for Topic 3

### **Innovative Algorithm Design: Designing and Prototyping New Algorithms for Neurology Applications**

**Name of Staff Member:**

Prof. Kostadin Mishev, Scientific Committee

**Host Institution:**

Jožef Stefan Institute (JSI), Ljubljana, Slovenia

**Exchange Period:**

12 May – 31 May 2025 (14 working days)

**Overview:** In alignment with Topic 3 of Deliverable D2.4, this Short-Term Staff Exchange focused on the design, development, and evaluation of cutting-edge retrieval-augmented algorithms to enhance data retrieval and knowledge representation for neurology applications. Given that the breadth and accuracy of retrieved information directly influence the quality and trustworthiness of large language models (LLMs) in medical contexts, a comprehensive analysis was conducted using a manually curated dataset. During the exchange at JSI, Prof. Mishev advanced the ChatMED project's data management platform by prototyping and benchmarking several retrieval-augmented generation (RAG) pipelines and by establishing best-practice protocols for fine-tuning LLMs in the medical domain.

**Objective Alignment with Topic 3:** The exchange fully addressed the objectives set out in D2.4 for Topic 3. A state-of-the-art survey of knowledge-representation and retrieval algorithms guided the selection of three RAG architectures: VectorRAG with hybrid search, GraphRAG built on SNOMED-CT ontologies, and a combined HybridRAG approach. Concurrently, a domain-specific evaluation dataset was constructed in collaboration with a medical expert, and a standardized protocol for medical-domain fine-tuning of LLMs was defined. These activities culminated in the creation of the RAGCare-QA benchmark, which supports rigorous assessment of RAG pipelines on 420

**Description of Work Performed:** During the first phase, Prof. Mishev conducted an in-depth review of contemporary tools and theoretical medical-knowledge questions, frameworks for knowledge representation and retrieval, identifying best practices and implementation strategies relevant to clinical applications. He then developed prototype implementations of three RAG pipelines: VectorRAG with hybrid dense-and-keyword search to balance precision and recall; GraphRAG leveraging structured SNOMED-CT knowledge graphs for context-aware responses; and HybridRAG, which fuses vector-based and graph-based retrieval to optimize both relevance and interpretability. In parallel, a domain-specific question set was curated, resulting in RAGCare-QA, a 420-item benchmark dataset. Finally, Prof. Mishev drafted a comprehensive protocol for fine-tuning LLMs in medical contexts, detailing data preparation, alignment techniques, and ethical considerations to ensure domain fidelity and responsible adaptation.

**Impact:** The STSE yielded three fully specified RAG prototypes and the open-access RAGCare-QA dataset, now available for community use. These resources establish a robust benchmarking framework for retrieval-augmented generation in medical education and decision support. The fine-tuning protocol provides a reproducible pathway for adapting LLMs to specialized clinical domains, thereby enhancing the reliability and explainability of AI-driven medical responses. Collectively, these outputs strengthen ChatMED's capacity to deliver accurate, context-aware, and trustworthy generative-AI tools in neurology.

**Conclusion:** Professor Mishev's exchange successfully met the Topic 3 objectives by delivering state-of-the-art RAG pipeline designs, a high-quality evaluation dataset, and a standardized LLM fine-tuning protocol tailored to the medical domain. These foundational contributions will be integrated into the ChatMED platform and serve as a springboard for future enhancements in algorithmic performance, domain alignment, and clinical validation.

**Outcome:** Submission of the manuscript “RAGCare-QA: A Benchmark Dataset for Evaluating Retrieval-Augmented Generation Pipelines in Theoretical Medical Knowledge” to *Data in Brief*.

## STSE Report for Topic 4

### Medical Software Development Life Cycle – Ensuring Quality and Adherence to Standards

**Name of Staff Member:**

Prof. Monika Simjanoska Misheva, Coordinator and Principal Investigator

**Host Institution:**

Jožef Stefan Institute (JSI), Ljubljana, Slovenia

**Exchange Period:**

12 May – 31 May 2025 (14 working days)

**Overview:** The Short-Term Staff Exchange (STSE) on Topic 4 was conducted at JSI under the heading “Medical Software Development Life Cycle – Ensuring Quality and Adherence to Standards,” originally scheduled for Year 2 but advanced to Year 1 to meet emerging project priorities and facilitate early collaboration with international partners. During her stay, Prof. Simjanoska Misheva spearheaded the alignment of ChatMED’s AI-enabled clinical software with both the EU AI Act and the MyHealth@EU interoperability framework, thus preemptively addressing high-risk classification requirements and lifecycle controls for clinical decision-support systems.

**Objective Alignment with Topic 4:** In accordance with Deliverable D2.4, Topic 4 mandates a systematic approach to quality assurance, regulatory compliance, and end-to-end lifecycle management for medical software incorporating AI components. This STSE fulfilled those requirements by producing a comprehensive tutorial entitled “A Step-by-Step Tutorial for Ensuring AI Act Compliance within the MyHealth@EU Framework.” The paper delivers a fully detailed engineering and compliance blueprint that maps the AI Act’s legal obligations transparency, risk management, documentation, and human oversight onto concrete software engineering artifacts aligned with ISO 13485, ISO 14971, ISO/IEC 42001, and MyHealth@EU interoperability standards. In doing so, it directly addresses the structured analysis of regulatory obligations, the mapping of those obligations across development phases, the definition of phase-specific compliance checkpoints, and the proposal of technical strategies particularly extensions to HL7 CDA and FHIR to integrate compliance seamlessly into clinical workflows.

**Description of Work Performed:** Prof. Simjanoska Misheva conducted a detailed review of the EU AI Act (Regulation EU 2024/1689) alongside the specifications of the MyHealth@EU interoperability framework. She identified all critical junctures in the software development lifecycle from project scoping and risk classification to post-market monitoring and re-conformity where compliance must be engineered by design. This analysis led to the creation of a phased compliance model with an accompanying, legally traceable tutorial for each stage of development. A comprehensive checklist was then formulated, linking every AI Act requirement to specific engineering tasks and deliverables, with annotations illustrating alignment to ISO



# Report on summer school

**Introduction:** Deliverable D2.4 originally envisaged Summer School 1 as an open, networking-centric forum to unite interdisciplinary researchers in artificial intelligence, neurology, software engineering, and healthcare policy. Its aim was to foster unconstrained knowledge exchange, ideation, and partnership formation to catalyse new collaborations. In practice, however, the ChatMED consortium's evolving priorities including the need to meet key performance indicators and to capitalize on momentum from intensive brainstorming sessions, comprehensive training activities, staff exchanges, and informal gatherings led to a recalibrated format. The revised approach preserves the foundational goals of networking and creative exchange but channels them toward a concrete objective: the co-development of a competitive proposal for the EIC Pathfinder Challenge "Generative-AI-Based Agents to Revolutionize Medical Diagnosis and Treatment of Cancer."

Accordingly, the official title for the inaugural summer school is: *Interdisciplinary Innovations: Encouraging Cross-Disciplinary Collaboration, Brainstorming, and Idea Generation through the EIC Pathfinder.*

This report documents the evolution from the original D2.4 vision into an intensive, proposal-co-creation laboratory. Building on D2.4 as the primary reference, the Summer School produced three additional deliverables:

1. A detailed presentation articulating the challenge's scope, specific objectives, expected outcomes, and projected impact.
2. An agenda structured around thematic workstreams, tailored to participants' professional expertise.
3. Daily summary notes capturing the ideas and insights generated during brainstorming sessions, organized according to the topic requirements.

Because these summary notes encapsulate the foundational concepts of the emerging proposal, they remain confidential until the application is formally submitted and evaluated by the funding authorities.

The summer school brought together a multidisciplinary cohort of experts, reflecting the core mission of ChatMED to integrate diverse perspectives from artificial intelligence, medicine, clinical practice, and compliance. The participation breakdown was as follows:

- **7 participants from the coordinating institution FCSE** covering AI architecture, data engineering, and language model design.
- **4 participants from the host JSI** with contributions to HPC-backed AI training pipelines.
- **3 participants from the FMN neurology team** providing expertise on diagnosis pathways and patient stratification strategies.

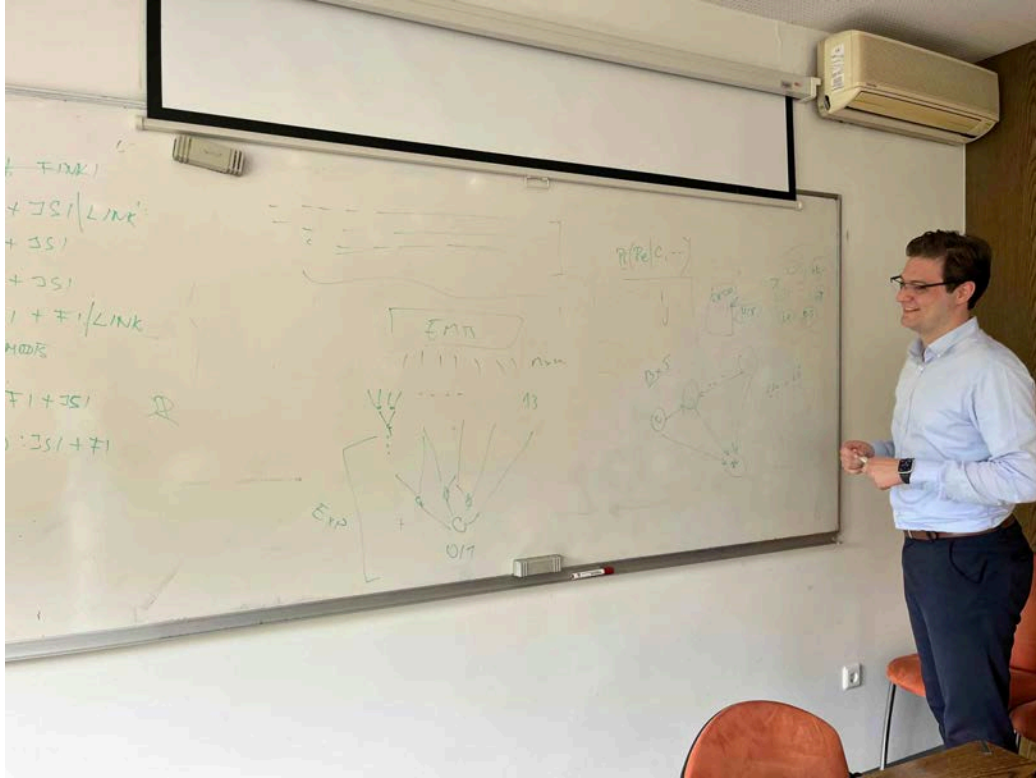
- **2 participants from TU Graz** leading the regulatory and quality assurance stream, particularly regarding compliance with the AI Act and MDR.

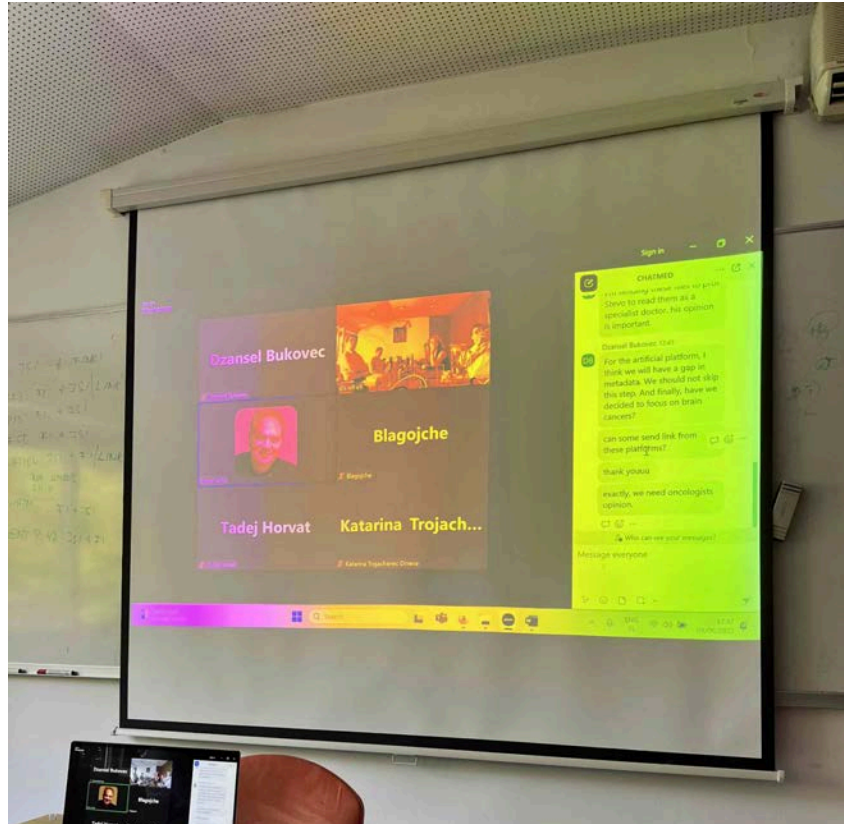
In addition to the core ChatMED partners, the team was strategically extended with representatives from Zan Mitrev Clinic in Skopje (ZMC), joining as a prospective partner for the upcoming EIC Pathfinder proposal. ZMC's involvement brought valuable insight into the realities of hospital-based diagnostics, data availability, and patient treatment workflows. Their active participation throughout the week enriched the brainstorming sessions, particularly in the areas of real-world data licensing, tumour staging via imaging, and treatment timelines.

This diverse and intentional participant mix helped to ensure that discussions were technically rigorous, clinically grounded, and regulatorily aware. More importantly, it served as a “warm-up session” for what is to come - the joint drafting, submission, and eventual implementation of a cutting-edge proposal under the Pathfinder Challenge. The environment during the school promoted openness, creative exchange, and shared ownership. Participants formed new working relationships that extended beyond institutional boundaries and laid the groundwork for deep, trust-based collaboration.

A few photos from the subsequent days showcase the energy, interaction, and collaborative spirit that defined the event.







## The Agenda design

To reflect the shift from open-ended networking toward a focused co-design of a future EIC Pathfinder proposal, the summer school adopted a **theme-based agenda**, aligning each day with a corresponding section of the targeted funding call. The structure preserved the interdisciplinary spirit of D2.4 but coupled it with concrete deliverable planning and role-specific discussion prompts.

Day	Daily focus	Daily keynote (morning briefing)	Workshop session					Informal events
			AI Team (FINKI, JSI)	Neurology Team (FMN)	Clinic Team (ZMC)	Genetics Team (ZMC)	Software Quality Team (TU Graz)	
Day 1	Introduction & Networking	<i>Understanding the Pathfinder Challenge scope and the motivation behind it</i>	What makes our proposed GenAI pipeline radically new and risky but plausible?	What are the current diagnostic and treatment bottlenecks for brain cancer patients? Where does diagnostic error, delay, or inconsistency most often occur?	What is the current diagnostic flow & limitations? Are there any standards we must follow (e.g., hospital protocols, GDPR constraints)?	What genetic/omics data types are currently used in brain cancer prognosis or therapy selection?	What criteria from EU AI Act and MDR, apply to the novel proposal?	Dinner
Day 2	Multimodal Fusion	<i>Drafting Excellence</i>	plan to fuse? How will we technically combine spatial, temporal, and semantic info?	Which data types (imaging, neurocognitive, clinical history) are critical for better prognosis?	What is the available hospital data? What is the sample size?	How can imaging/clinical data be linked with molecular insights?	What metrics will we use for robustness, fairness, and explainability?	Networking
Day 3	Data Challenges	<i>Open Science and Data</i>	Science practices? What methods can be used to create highly realistic synthetic data?	Propose synthetic scenarios for validating the proposed concept.	Which public and hospital datasets do we expect to use or augment? Are licenses/ethics in place?	Identify relevant genomic/epigenomic data sources.	Discuss reproducibility protocol.	Networking
Day 4	Explainability & Trustworthy AI	<i>Understanding AI robustness</i>	How will we address interpretability and explainability?	Define what interpretability means in neuro-oncology. Can GenAI realistically support early diagnosis, decision-making, or patient stratification in our context?	How can clinicians interpret GenAI outputs? What will increase their trust and adoption?	Discuss genomic data interpretability in clinical settings.	Plan for validation in clinical setting.	Networking
Day 5	From Research to Impact	<i>Writing impact</i>	What will be the outcomes (platforms, benchmarks, datasets, metadata, trace logs, audit trails)?	Define benefits to clinicians, patients.	Highlight cost/time benefits to hospital workflows. What outcomes will we compare: diagnostic accuracy, time-to-treatment, risk stratification?	Show biomarker discovery potential and personalized treatment.	Estimate performance & safety in clinical-grade SW.	Networking
Day 6	Elaborating Implementation Plan	<i>Writing work plan</i>	Draft WP for AI.	Draft WP for Neurology Knowledge.	Draft WP for Clinical Trial.	Draft WP for Genetic Data Pipeline.	Draft WP for compliance with MDR & AI Act.	Networking
Day 7	Showcase and Wrap-Up	<i>Define KPIs</i>	deliverables and publications (define working titles).	Estimate number of publications (define working titles).	Estimate number of publications (define working titles).	Estimate number of publications (define working titles).	Estimate number of publications (define working titles).	Farewell coffee

Each participant group, AI (FINKI, JSI), Neurology (FMN), Clinical (ZMC), Genetics (ZMC), and Software Quality & Compliance (TU Graz), was invited to contribute to a shared narrative structured across seven thematic days.

In addition to the structured workshops, each day included:

- A **morning keynote briefing**, summarizing the focus area and aligning goals across domains.
- An **informal networking session**, facilitating lateral idea exchange and informal mentoring.

## Daily Implementation and Outcomes

The seven-day summer school was organized around evolving six core innovation areas derived from the EIC Pathfinder Challenge theme. Each area was advanced iteratively through collaborative ideation, resulting in clearly traceable novel contributions. Below is a synthesized mapping of how each target area evolved across the days and the key outcomes:

### A) GenAI Tools for Multimodal Integration

- Day 1: Mapped limitations in current GenAI pipelines (e.g., static vs. temporal modalities).

- Day 2: Agreed to add a temporal dimension to the knowledge graph and integrate real-time clinical events.
- Day 3: Anchored data in a Data Vault modeling for consistent multimodal handling.
- Day 4: Integrated LLM + XGBoost pipelines to enable explainable decision trees.
- Day 5-6: Aligned pipeline adaptation to patient imaging modality availability.
- Outcome: A temporal-spatial fusion architecture with an explainability layer built-in.

## B) Medical Data Augmentation

- Day 1: Proposed mining of case-report literature for model training.
- Day 2: Defined a synthetic data generation pipeline, especially for rare tumour types.
- Day 3: Integrated synthetic schema with the Data Vault.
- Day 4: Need to create and test synthetic question-answer pairs using clinical prompts.
- Day 5-6: Identified data scarcity in CT and PET imaging to prioritize augmentation.

Outcome: Data augmentation pipeline anchored in reproducibility and rarity management.

## C) Medical Knowledge Representation & Integration

- Day 1: Introduced prime implicant logic for inference traceability.
- Day 2: Flagged the Human Brain Project (HBP) 3D atlas as a semantic map resource.
- Day 3: Wrapped HBP as a plug-in and committed to ontology extension.
- Day 4: Developing a decision tree with semantic traceability.
- Day 5-6: Highlighted the lack of automated slice selection, adding it as a pipeline module.

Outcome: Unified KG integrating brain maps, guidelines, and patient trajectories.

## D) External Data & Collaborations

- Day 1: No clear external linkages yet.
- Day 2: Performed a landscape scan of EU projects, particularly HBP.
- Day 3: Acquiring access to 4TB of imaging data from HBP.
- Day 4: Aligned with OpenAI HealthBench as an external benchmark.
- Day 5-6: Created an ethics/license checklist for partner data.

Outcome: Strategic data partnerships activated and data use protocols initiated.

## E) Predictive Diagnosis & Personalized Treatment

- Day 1: Defined the need for a front-door triage agent and therapy output mapping.
- Day 2: Added an early-detection flow, confirmed high-risk/high-gain value.
- Day 3: Selected CT staging as a pilot.
- Day 4: Ensured treatment outputs are validated by XGBoost safeguard agents.
- Day 5-6: Confirmed imaging-genetics links for tumour subtyping.

Outcome: Full prototype sketch for LLM-supported therapy refinement.

#### F) Evaluation, Robustness & Interpretability

- Day 1: Agreed robustness must be measurable and comparative.
- Day 2: Introduced one-hop traceability for outputs and started logging schema.
- Day 3: Built initial LLM-to-LLM messaging protocol to enforce causality.
- Day 4: Defined a two-phase validation using HealthBench + clinician A/B tests.
- Day 5: Discussed trade-off between model accuracy and explainability in neuroimaging.
- Day 6: Integrated Topological Data Analysis (TDA) and causal graph visualization.

Outcome: Multi-layered evaluation stack combining regulatory, clinical, and technical KPIs.

Day 7 was reserved for final polishing of the conclusions.

This continuous, structured progression across target areas will make it possible to turn abstract ambitions into operational modules in the upcoming project proposal with deadline on 29th of October. Each day's contributions built logically on the previous, ensuring coherence and incremental innovation aligned with both Pathfinder ambitions and the AI Act high-risk system criteria.

The first ChatMED Summer School evolved into a highly productive and strategically aligned event that bridged the original networking intentions of D2.4 with a concrete roadmap for high-impact proposal development. Across seven thematic days, participants advanced six innovation tracks from abstract conceptualization to actionable components, each mapped to real-world constraints in clinical practice, data governance, AI regulation, and scientific feasibility.

### Synthesis of Achievements

- A **multi-agent GenAI architecture** was drafted, incorporating multimodal data fusion, temporal extensions, and explainability safeguards.
- A **synthetic data pipeline** was outlined, and aligned with reproducibility goals.
- Participants agreed on a **unified knowledge representation**, ontology plug-ins, and logic-based traceability.
- New **external data partnerships** were activated.
- A **front-door triage agent** and **LLM-assisted therapy recommender** were mapped out, with validation tasks delegated to XGBoost decision trees.
- A **robust evaluation framework** was drafted combining HealthBench, clinical user studies, and causal topological analysis.



in neurology), Topic 2 (vision-enabled LLMs for multimodal neuroimaging), Topic 3 (prototyping Retrieval-Augmented Generation pipelines), and Topic 4 (medical-software lifecycle compliance with EU standards). Each exchange delivered architecture designs, benchmark datasets (including the open-access RAGCare-QA), prototype agents, compliance blueprints, and manuscripts in preparation.

### 3. Summer School

The first Summer School retitled “*Interdisciplinary Innovations: Encouraging cross-discipline collaborations, brainstorming, and idea generation through EIC Pathfinder*” reoriented the original networking format into a week-long proposal lab for an EIC Pathfinder Challenge on generative-AI agents for cancer diagnosis and treatment. Participants co-created a detailed presentation, agenda, and confidential daily summaries across six innovation tracks (multimodal fusion, data augmentation, knowledge representation, external partnerships, predictive diagnosis, evaluation), forming the foundation for an October 2025 proposal.

*The skills which was boosted from STSE, summer school and training sessions:*

#### Technical and AI-Related Skills:

1. Generative AI (GenAI) Pipeline Development Improving multimodal integration, temporal data handling, and explainability layers.
2. Data Engineering and Modeling Implementing Data Vault modeling for structured multimodal data handling.
3. Machine Learning and AI Training Integrating LLM + XGBoost pipelines for explainable decision trees.
4. Synthetic Data Generation Developing pipelines for rare tumour types and reproducibility.
5. Knowledge Representation and Ontology Management Extending medical ontologies and integrating semantic maps (e.g., Human Brain Project 3D atlas).
6. Topological Data Analysis (TDA) and Causal Graph Visualization Enhancing interpretability in AI-driven diagnostics.

#### Medical and Clinical Skills:

7. Medical Imaging Analysis Addressing CT/PET imaging scarcity and improving tumour staging.
8. Patient Stratification and Diagnosis Pathways Refining early-detection flows and therapy output mapping.
9. Clinical Data Augmentation Mining case-report literature and generating synthetic Q&A pairs for training.

#### Regulatory and Compliance Skills:

10. AI Act and Medical Device Regulation (MDR) Compliance Ensuring high-risk AI system adherence.
11. Ethics and Data Licensing Developing checklists for partner data usage and real-world data governance.

### Collaborative and Proposal Development Skills:

12. Interdisciplinary Brainstorming Bridging AI, neurology, clinical practice, and policy.
13. Strategic Partnership Building Engaging external collaborators (e.g., Zan Mitrev Clinic, Human Brain Project).
14. Proposal Writing and Grant Strategy – Structuring a competitive EIC Pathfinder submission.
15. Project Coordination & Role-Specific Contributions – Aligning expertise across institutions (AI, HPC, neurology, compliance).

### Soft and Organizational Skills:

16. Creative Problem-Solving – Iterative ideation and structured innovation.
17. Cross-Disciplinary Communication Facilitating informal mentoring and lateral idea exchange.
18. Trust-Based Collaboration Forming new working relationships beyond institutional boundaries.