Horizon Europe (HORIZON)

HORIZON-WIDERA-2023-ACCESS-02
Twinning

Progress Report

Progress Report

COVER PAGE

PROJECT	
Project number:	101159214
Project acronym:	ChatMED
Project name:	Bridging Research Institutions to Catalyze Generative AI Adoption by the Health Sector in the Widening Countries
Call:	HORIZON-WIDERA-2023-ACCESS-02: Twinning
Type of action:	CSA
Responsible service:	REA.C3
Project starting date:	01/09/2024
Project duration:	36
Period covered by the report:	from 01/09/2024 to 31/12/2025

1. EXPLANATION OF THE WORK CARRIED OUT AND OVERVIEW OF THE PROGRESS

1.1 Executive Summary

The following Table 1 lists the specific objectives for the project as described in Section 1.1 of Annex I (Part B) of the Description of the Action along with the measurable outcome and verification as planned vs. the actual results achieved by 31.12.2025.

Table 1. Specific objectives matched with the mid-project results outcomes.

Specific Objectives	**Relation to the programme**	**Measurable outcome and verification**	**Mid-project outcomes**
SO1: *Establish a Comprehensive Generative AI Training and Knowledge Exchange Hub.*	Improved excellence capacity and resources in Widening countries enabling to close the still apparent research and innovation gap within the EU.	Establish a SOTA training facility dedicated to GenAI in healthcare with a specialized focus on multimodal data integration for neurology. *Verification*: D2.1, D2.2, D2.3, D2,4, D3.1, D1.1.1.	1. Equipment with H100 GPU procured at the coordinator's facility. 2. AI Antenna established in connection with the Greek Pharos AI Factory (Grant received in October 2025). 3. DMP plan with an architecture for scientific data management platform established (D2.1). 4. Detailed communication, dissemination and exploitation plan established with tables to track the annual progress (D2.2). 5. Detailed training curriculum, STSEs and summer school programs are established with the focus on AI in neurology and mastering the project management (D2.4). 6. Benchmarking of the coordinator's organizational structure with the leading universities (D3.1). 7. Establishment of project website and communication social and open science channels to keep all the activities (D2.3).
SO2: *Foster Innovation by Promoting the Exchange of Creative Research Methodologies and Strategies.*	Enhanced strategic networking activities between the research institutions of the Widening countries and two leading counterparts at EU level.	Organize and conduct brainstorming sessions, lectures, workshops, conferences, attend conference events, publish joint papers, organize special journal issues, and initiate interdisciplinary research collaborations. *Verification*: D1.1.1,	1. Organized brainstorming session in September 2024. 2. Organized a special workshop at IS 2024. 3. Organized art exhibition in November 2024. 4. Organized message transfer activity in November 2024. 5. Organized training session in April 2025. 6. Organized industry panel in April 2025. 7. Organized training session in July 2025.

		D1.1.2, D2.2, D2.3, D2.4, D4.1, D5.1, D6.1, D8.3.	8. Organized training session in October 2025. 9. Organized a special workshop at IS 2025. 10. Initiated and submitted project proposals with 13 new partners at HORIZON-HLTH-2025-01-CARE-01. 11. Initiated and submitted project proposal with 7 new partners at HORIZON-EIC-2025-PATHFINDERCHALLENGES-01-02. 12. Accepted a consortium invitation to participate in a project proposal submitted to HORIZON-CL4-2025-03. 13. Accepted a consortium invitation to participate in a COST action for the call OC-2025-1-28842. 14. Submitted bilateral project between North Macedonia – Germany. 15. Submitted bilateral project between North Macedonia – Austria. 16. Signed memorandums for collaborations with four industry partners. 17. Published 8/10 impact factors journals. 18. Published 17/30 conference papers. 19. Networking as a panelist at the 4th AmCham North Macedonia Healthcare Conference. 20. Networking for Future Healthcare Innovation at APHP & Inserm in Paris.
SO3: *Advance Researcher Mobility and Expertise Sharing.*	Raised reputation, research profile and attractiveness of the coordinating institution from the Widening country and the research profile of its staff.	Launch the STSE programs, organize summer schools for the PhD students and the senior researchers. *Verification*: D2.4, D4.1, D5.1, D6.1.	1. Launched the first round of STSE activities planned for year 1 in May 2025. Five FCSE staff members were sent to JSI. 2. Organized first summer school at JSI in June 2025.
SO4: *Standardize and*	Strengthened research management	Publish the training curriculum, organize research and	1. Training curriculum as well as the other deliverables have been openly published at the project website and at Zenodo.

Disseminate Best Practices in Research and Administration.	capacities and administrative skills of the staff working in institutions from the Widening countries.	administration training activities, summarize the best practices. *Verification*: D2.4, D4.1, D5.1, D6.1, D1.1.1, D1.1.2.	2. Training sessions on research and administration have been organized with highlights sublimated in form of raw materials available at the project website and as blogs. Shared also on the social channels.
SO5: *Create Multimodal Generative ChatMED Platform.*	Development of new approaches in Research and Innovation collaboration.	Develop a multimodal ChatMED platform designed for neurology that integrates diverse diagnostic and analytical tools and measures. *Verification*: D1.1.2, D2.5, D7.1, D7.2, D8.1, D8.2, D8.3.	1. ChatMED HITL platform established and available as an open source code. 2. ChatMED platform established as a nested multi-agent ecosystem covering distinct modalities as separate multi-agent systems. 3. Pharma-domain multi-agent system published in IF journal and available as open source code. 4. Vision-enabled LLMs evaluated and a comprehensive paper submitted for review. Currently available as a preprint at Zenodo. 5. Bioinformatics LLMs landscape critically reviewed and submitted for evaluation. 6. Policy aspects of ChatMED investigated and published.

1.2 Explanation of the work carried out per WP

1.2.0 Research component

The research component has already produced a substantial body of peer-reviewed scientific output, significantly exceeding expectations for the reporting period. FCSE leads the majority of research activities, in line with the requirement that at least 50% of the research budget is allocated to the coordinator, with JSI, FMN, and TU Graz contributing to specialised evaluation, validation, and methodological tasks.

Table 2 summarises the unique research outputs produced within the research component, grouped by maturity level. Partner columns indicate involvement in each output, while the 'Unique outputs' column reflects the total number of distinct results.

Table 2. Unique research outputs by maturity and beneficiary involvement.

Output category	Status	Unique outputs	FCSE	JSI	FMN	TU Graz
Journal papers	Published	**8**	7	3	1	0
	Under review	**3**	2	0	0	1
	In progress	**4**	3	0	1	0
Datasets	Published	**1**	1	1	0	0
	In progress	**1**	1	1	0	0

Conference papers	Published	**18**	6	4	5	3

1.2.1 *Work Package 1: Project Management*

WP1 ensures the effective coordination, monitoring, quality assurance, financial control, and risk management of the ChatMED project. The work package provides the managerial backbone of the project, enabling timely execution of technical activities, transparent performance tracking, and alignment with scientific, ethical, and exploitation objectives.

T1.1 Monitoring and reporting. Since the start of the project, the coordinator (FCSE) has established a shared project management and monitoring space enabling transparent tracking of all activities, partner contributions, and KPIs. This shared workspace at Google Drive supports continuous reporting, version control, and progress monitoring across work packages. Table 3 presents the project KPIs as monitored through all the predefined categories, considering only the work that is completed (published papers/datasets, submitted proposals, signed collaborations).

Table 3. Key performance indicators (unique outputs and partner involvement).

KPI category	Unique outputs	FCSE	JSI	FMN	TU Graz
Journal papers	**8**	7	3	1	0
Conference papers	**18**	6	4	5	3
Datasets	**1**	1	1	0	0
Project proposals	**7**	7	2	1	3
Industry memorandums	**4**	4	0	0	0

Effective consortium coordination has been ensured through regular project meetings, resolving inter-work-package dependencies, and reviewing emerging risks. The project foresees six **consortium meetings** over its duration. By the end of the reporting period, four meetings had been successfully completed:

- Kick-off meeting on 24.09.2024 (online)
- 1st consortium meeting on 13.11.2024 (hybrid)
- 2nd consortium meeting on 26.04.2025 (on-site, Strumica)
- 3rd consortium meeting on 09.10.2025 (on-site, Ljubljana, with TU Graz participating online)

All meetings were formally minuted and, where appropriate, recorded. Action points and decisions were systematically documented and followed up, ensuring traceability of technical decisions and timely mitigation of identified issues.

Deliverable progress has been closely monitored as part of T1.1 to ensure timely submission and quality consistency. By 31.12.2025, nine out of eighteen deliverables had been submitted on time. All submitted deliverables followed a uniform project template, were uploaded to the reporting system in their final or appropriate interim versions, and, where designated as public (PU), were published on the project website and in the Zenodo ChatMED community.

The project defines four **milestones**. By the end of the reporting period, one milestone had been completed on time, and one milestone is pending with a deadline of 31.01.2026. The completed **milestone MS1** established the foundational framework for the project's training, visibility, and research activities and encompassed the following deliverables:

- D2.1 Data management plan
- D2.2 Plan for communication, dissemination, and exploitation
- D2.3 Website establishment

- D2.4 Detailed training curriculum and talent mobility programmes
- D2.5 Report on human evaluation framework created for benchmarking.

Together, these deliverables provide the methodological, operational, and governance foundations for all downstream technical work packages, enabling structured research execution, human-in-the-loop evaluation, and coordinated dissemination and exploitation activities. Each is separately explained in the corresponding work package.

T1.2 Quality assurance. Quality assurance activities have been implemented consistently across all project phases to ensure the reliability, comparability, and scientific robustness of project outputs. These activities are designed to support both internal quality control and external evaluability of results, in line with the project's objectives and stakeholder expectations.

Quality assurance is operationalised through a set of systematic measures, including the use of harmonised deliverable templates across all work packages, ensuring consistency in structure, level of detail, and reporting standards. In addition, all deliverables undergo internal peer review prior to submission, allowing for early identification and correction of issues related to clarity, completeness, methodological soundness, and alignment with contractual requirements. Finally, quality control is reinforced through explicit alignment with predefined evaluation dimensions, ensuring that outputs can be assessed coherently across technical, clinical, and user-centred perspectives.

Given the rapid evolution and increasing heterogeneity of large language models, the project has placed particular emphasis on the development of benchmarking datasets according to the highest ethical, legal, and quality standards, therefore enabling reproducible and comparative evaluation. To this end, the project committed to the development of benchmarking datasets aligned with the human evaluation framework defined in Deliverable D2.5. These benchmarks are designed to support multi-dimensional assessment of LLM performance across key application areas relevant to the project's objectives, including:

- Technical Reliability (Testing robustness, feasibility, and uncertainty)
- Summarization of Patient History
- Medical Knowledge Inquiry (Neurological Conditions)
- Prediction Tasks (e.g., Neurological Diagnosis, Treatment Recommendations)
- Administrative Support Tasks (e.g., Autocompleting Doctors' Reports)
- Patient Educational Support

Several of these benchmarking datasets have already been partially published, providing early validation assets for downstream research activities. Performance evaluations derived from these benchmarks will be disseminated through peer-reviewed journals and conference publications, contributing to both the scientific visibility of the project and the transparency of its evaluation methodology.

T1.3 Financial management. The project operates under a lump-sum funding model, in accordance with Horizon Europe regulations. Financial management under T1.3 focuses on ensuring sound internal control, transparency, and readiness for potential audits, while avoiding unnecessary administrative burden on beneficiaries.

As coordinator, FCSE maintains internal records of personnel effort in the form of timesheets as supporting documentation. These records are used solely for internal monitoring, resource planning, and risk mitigation purposes and do not constitute formal financial reporting under the lump-sum scheme.

Financial management for the other beneficiaries is carried out locally at each partner institution, in line with their internal procedures and without centralised financial oversight by the coordinator. This approach is fully aligned with Horizon Europe practices for lump-sum projects, where financial justification is based on task and deliverable completion rather than on detailed cost reporting.

Following the receipt of pre-financing, the coordinator redistributed the corresponding shares to the

beneficiaries in accordance with the Consortium Agreement. This redistribution supports the timely execution of project activities and partner liquidity, while remaining fully compliant with the principles of the lump-sum funding model.

T1.4 Risk management. Risk management under WP1 has been implemented as a continuous and proactive process aimed at identifying, assessing, and mitigating risks that could affect the timely and successful execution of the project. Risks initially identified at the proposal stage have been systematically monitored throughout the reporting period, while additional unforeseen risks emerging during project implementation have been documented and addressed through appropriate corrective measures.

1.2.2 *Work Package 2: Establishing Coordinating and Supporting Measures*

WP2 represents the structural backbone of the ChatMED project. It establishes the coordination, governance, and support mechanisms required to ensure effective project execution, clear allocation of responsibilities, and long-term sustainability beyond the project lifetime. As the Widening coordinator, FCSE leads this work package with the objective of building organizational awareness, strengthening governance structures, and embedding durable processes that remain applicable after the project concludes.

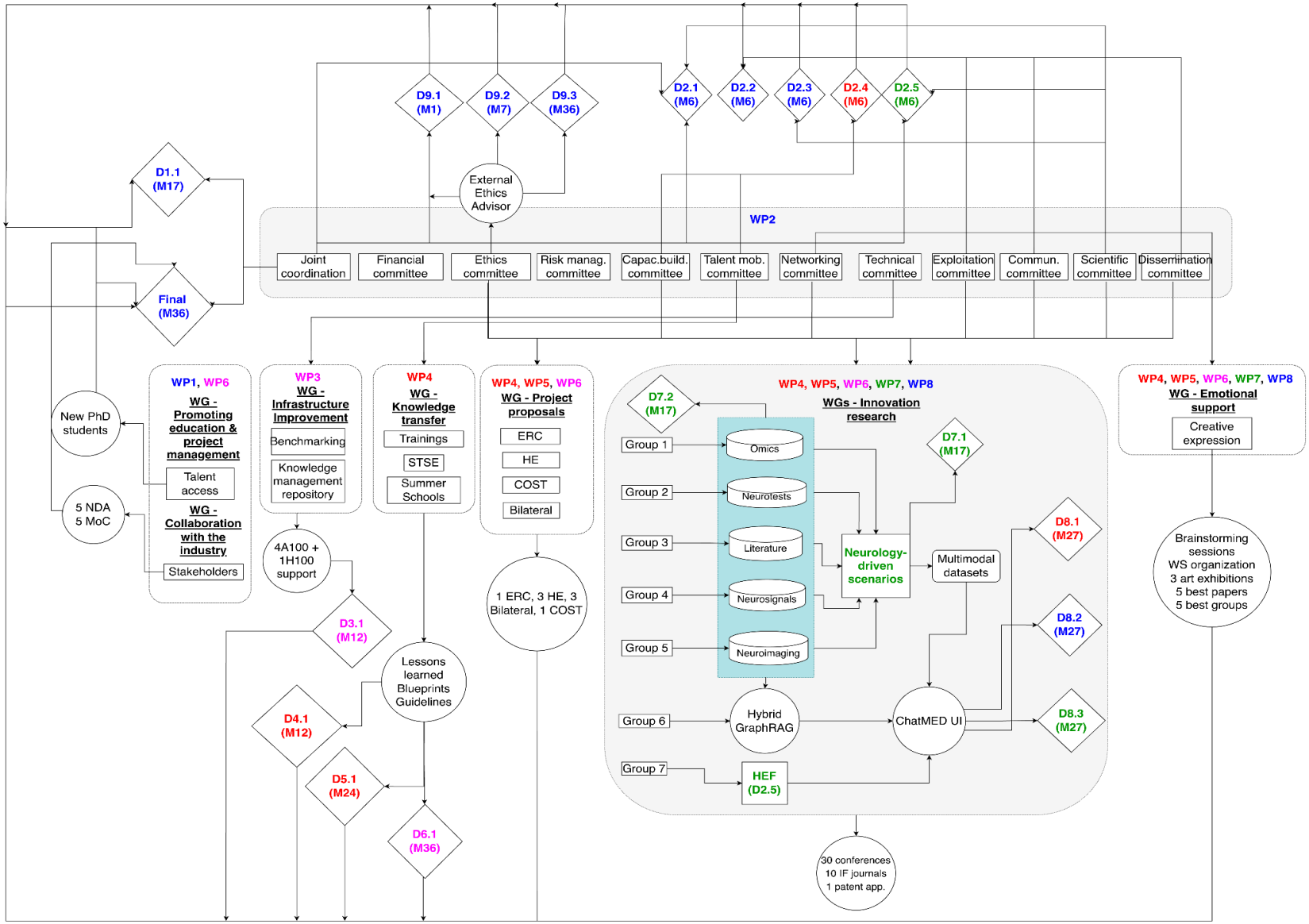
The deliverables produced under WP2 form a lasting framework for coordination, training, data governance, and human oversight. In light of the entry into force of the EU AI Act, several WP2 outputs, most notably the Data Management Plan and the Human Evaluation Framework, are directly aligned with emerging regulatory requirements related to data governance, transparency, and human oversight. Each task under WP2 contributes to establishing these foundations, as described below.

T2.1 Establishment of coordinating and supporting bodies. This task was initiated immediately at the start of the project. The coordinator (FCSE) designed and implemented a governance scheme with clearly defined roles and responsibilities for all coordination and advisory bodies foreseen in the project proposal. This included the formal establishment of committees and the creation of thematic working groups aligned with the project's technical, training, dissemination, and ethical objectives.

An organizational structure was defined, as presented in Figure 1 (zoom in allows for more clear details), and shared with all beneficiaries, mapping committees and working groups to specific project activities and responsibilities. All coordinating bodies and working groups were consolidated in a shared project file, which serves as a living reference linking entities, roles, and task ownership across the project.

The governance structure and participation scheme were formally presented during the kick-off meeting. All beneficiaries received explicit invitations to join the coordinating bodies and relevant working groups, ensuring early engagement, transparency, and shared ownership of project activities. This early establishment of governance mechanisms enabled effective coordination, facilitated communication across work packages, and provided a stable framework for decision-making throughout the reporting period.

Figure 1. Organizational structure of the ChatMED project.



T2.2 Establishment of the Data Management Plan. Under T2.2, the consortium developed and delivered the Data Management Plan (D2.1), which goes beyond a standard Horizon Europe DMP and is positioned as a strategic coordination, compliance, and sustainability instrument. The DMP defines how data is collected, processed, governed, shared, and preserved across the project, in full alignment with FAIR principles, GDPR requirements, and ethical guidelines.

A central element of the DMP is the **TriZ-Flow architecture**, a three-zone data processing pipeline (Bronze, Silver, Gold) designed to support privacy-preserving data ingestion, standardization, validation, and transformation. This architecture operationalizes key requirements of the EU AI Act related to data governance, traceability, data quality, and risk mitigation, while simultaneously enabling reproducible research workflows and structured human oversight. The TriZ-Flow design conceptually underpins the Human Evaluation Framework (D2.5), providing the methodological basis for governing any newly collected, potentially sensitive data used for LLM training and evaluation, distinct from publicly available sources, with a focus on control, auditability, and ethical management.

Importantly, the DMP and the TriZ-Flow architecture are explicitly conceived as **long-term assets** extending beyond the ChatMED project. They serve not only as internal coordination tools but also as reusable infrastructure and methodological references for future projects led by the Widening coordinator.

As a direct follow-up to discussions held on 9 January 2026 with the national health system vendor, and building on issues raised during the 4th AmCham Conference where ChatMED participated as a panelist, the TriZ-Flow architecture is being positioned as a mechanism for transforming national health data into ready-to-use, fully anonymized, standardized datasets for research purposes which are foreseen to be deployed as an Open Data catalogue, enabling compliant reuse of health data for research, innovation, and policy development, while maintaining strict privacy and regulatory safeguards.

Through this approach, T2.2 contributes not only to project compliance and quality assurance but also to national and regional capacity building, aligning ChatMED with broader European objectives on trustworthy AI, open science, and sustainable data ecosystems.

T2.3 Plan for communication, networking, dissemination & exploitation (CNDE). Under Task T2.3, the consortium developed and delivered **D2.2 – Plan for Dissemination, Communication, and Exploitation**, which defines the strategic framework for all CNDE activities throughout the project lifecycle and beyond. D2.2 establishes a coherent and integrated CNDE strategy, structured around clearly defined objectives, target audiences, channels, timelines, and key performance indicators. The deliverable defines:

- Dissemination pathways for scientific outputs, including journals, conferences, benchmarks, and datasets.
- Communication activities targeting academia, healthcare professionals, policymakers, industry, and the general public.
- Exploitation mechanisms aimed at sustainability, technology transfer, policy alignment, and follow-up funding.

Although Task T2.3 formally concludes at Month 6 with the delivery of D2.2, CNDE activities are implemented as transversal and living processes throughout the project. These activities are embedded as dedicated tasks within WP4, WP5, WP6, and WP8, covering the entire project duration and ensuring continuous alignment with technical, training, and research developments. The outcomes of these CNDE activities are systematically captured and reported through the corresponding work package deliverables, in particular D4.1, D5.1, and D6.1, which provide consolidated evidence of implementation, impact, and achieved KPIs.
D2.2 is designed as a living document and foresees at least one update before the end of the project, with the final version extending CNDE planning up to four years beyond project completion, thereby supporting long-term impact and institutional sustainability.

Quantitative KPIs and timelines available in the annexes of the D2.2 enable systematic monitoring and reporting. The following tables 4, 5, and 6 quantify the CNDE achievements up to 31.12.2025.

Table 4. Mid-project summary table of actual realization of the dissemination plan.

Activity	Year 1				Year 2	
	Q1	Q2	Q3	Q4	Q1	Q2 (only Dec 2025)
Online webpage posts	N/A	37	4	5	1	21
Conference publications	10				8	
Journal articles	5				3	
Training sessions			1	1	1	
Summer schools				1		
Last update: 31.12.2025						

Table 5. Mid-project summary table of actual realization of the communication plan.

Activity	Year 1				Year 2	
	Q1	Q2	Q3	Q4	Q1	Q2 (only Dec 2025)
Message transfer events	1					
Social networks posts	15				9	
Short-term staff exchange			5			
Consortium meetings	2		1		1	
Brainstorming sessions (informal meetings)	1	5+	10+	10+	10+	4
Art exhibitions	1					
Best paper awards	2				1	
Best working group awards	N/A				TBD	
Press releases	1					
Newsletters	4				2	
Meetings with policy makers/decision makers	1	1	1	3	2	1
Last update: 31.12.2025						

Table 6. Mid-project summary table of actual realization of the exploitation plan.

Activity	Year 1				Year 2	
	Q1	Q2	Q3	Q4	Q1	Q2 (only Dec 2025)
Trademarking		1				
Patenting analysis	N/A				N/A	
Integrated solutions with industry or healthcare services	1					
NDAs signed					4	
Collaboration memorandums					4	
Exhibitions and trade fairs	N/A				N/A	
Horizon Europe follow-ups					4	
COST					1	
ERC						
Bilateral	1				1	
Last update: 31.12.2025						

As part of Task T2.3, **Deliverable D2.3 – Website Establishment** was completed and submitted as a public deliverable (PU), providing the central digital platform for communication, dissemination, and transparency of the ChatMED project. The project website (https://chatmed-project.eu/) is fully operational and regularly updated, as well as the established social channels at LinkedIn, X, Instagram, Threads, and Facebook. By the end of the reporting period, the website hosts 19 blog posts, 13 event announcements and reports, 15 news items, and 6 newsletters (five quarterly editions and one New Year's edition). In addition, the Knowledge Repository contains 27 curated items, including presentations, reports, and public outputs from project activities. Dedicated sections provide structured access to all project deliverables and milestones, ensuring transparency and traceability of progress. Web analytics for the last 90 days indicate 690 unique visitors, representing a 117.7% increase compared to the previous period, with traffic dominated by direct access (80.8%) and organic search (16.1%). The homepage, knowledge repository, and "About" section are the most frequently accessed pages, demonstrating effective visibility of core project information and research outputs. These indicators confirm that the website functions as an active dissemination hub supporting CNDE objectives and providing measurable outreach to diverse stakeholders.

T2.4 Training curriculum development. Under Task T2.4, the consortium developed and delivered Deliverable D2.4 – Detailed Training Curriculum and Talent Mobility Programme, which defines a structured, modular training framework supporting capacity building, knowledge transfer, and long-term skills development within the ChatMED project. The curriculum addresses the needs of diverse target groups, including early-stage researchers, senior researchers, clinicians, and technical staff, and is aligned with the project's research, evaluation, and deployment objectives. The training curriculum is organised around thematic modules covering generative AI foundations, large language models in healthcare, regulatory and ethical considerations, data governance, benchmarking and evaluation methodologies, and applied use cases in neurology. It integrates multiple training formats, including lectures, hands-on workshops, short-term scientific exchanges (STSEs), and summer school activities, enabling both theoretical understanding and practical skill acquisition. Importantly, the curriculum is designed as a living training framework, allowing iterative updates as new models, tools, and regulatory requirements emerge. This ensures that the

training activities remain relevant throughout the project duration and beyond, supporting long-term sustainability and institutional capacity building at the Widening coordinator and partner institutions. The curriculum also provides the foundation for training-related activities implemented and reported in subsequent work packages.

T2.5 Establishing the Human Evaluation Framework. During the first year of the ChatMED project, the consortium was confronted with the rapid and continuous release of new LLMs demonstrating strong potential for application in the neurology domain. To enable systematic, reproducible, and clinically meaningful assessment of these models from the perspective of medical experts, Task T2.5 focused on the design and implementation of a Human Evaluation Framework (HEF), comprehensively described in D2.5. The HEF is grounded in a **human-in-the-loop** paradigm and provides a structured methodology for evaluating LLM outputs across multiple dimensions relevant to healthcare, including technical performance, quality and reliability, safety and bias, security and privacy, and acceptance. To operationalise the framework, a dedicated ChatMED Human Evaluation Platform (ChatMED-HITL[12]) was developed. The platform was designed to be intuitive and accessible for clinicians, while remaining technically robust and extensible.

The ChatMED-HITL platform allows evaluators to upload JSON files containing benchmark questions and corresponding LLM-generated answers. Evaluations are organised into workspaces, as presented in Figure 2, each comprising a defined pool of benchmark questions, a pool of LLMs under evaluation, the concrete model-generated answers under review, a predefined set of expert evaluation metrics, and a free-text comment field enabling nuanced qualitative feedback. This structure enables systematic collection, aggregation, and analysis of expert assessments, supporting evidence-based comparison of models across tasks and modalities.

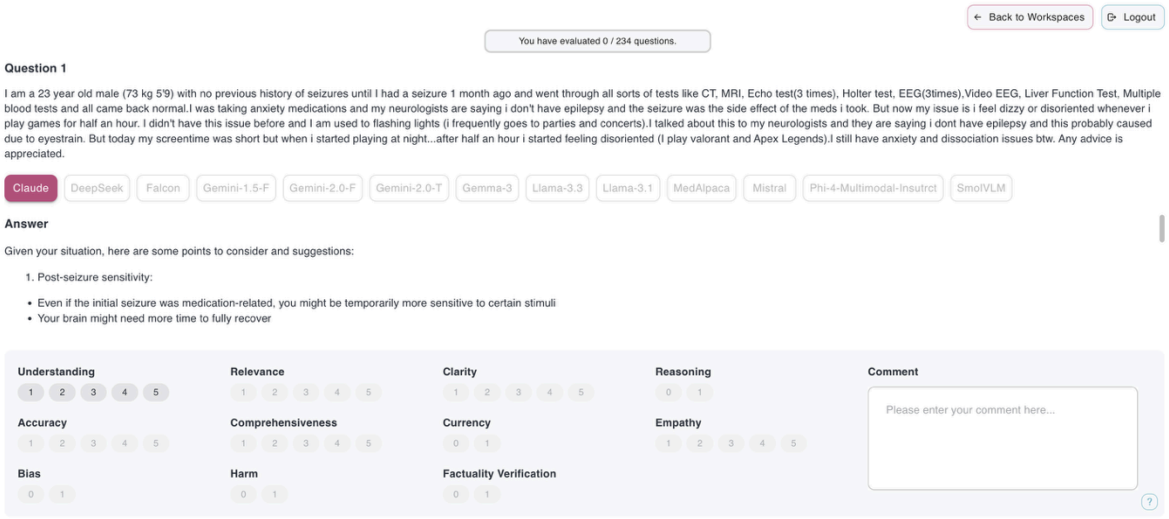


Figure 2. ChatMED Human Evaluation Platform.

Through this implementation, the project established a controlled and scalable mechanism for collecting expert feedback, enabling informed selection of suitable foundation models for subsequent fine-tuning and downstream task development. To date, the platform has been applied to the patient education support downstream task, where a benchmarking dataset of 234 questions has been evaluated across 13 candidate LLMs, providing early comparative insights into model suitability.

In addition to the static evaluation metrics defined in D2.5, the consortium developed an **open-source Python library**[3] that supports LLM-assisted suggestions of domain-specific evaluation metrics. This library is designed to be added as an extension to the ChatMED-HITL platform,

[1] https://github.com/ChatMED/hef
[2] https://github.com/ChatMED/hef-frontend
[3] https://github.com/ChatMED/hef-metrics-generator

enabling adaptive refinement of evaluation criteria as new tasks and clinical requirements emerge.

By supporting informed model selection, documenting expert judgement, and enabling intervention before further fine-tuning or deployment, the platform aligns with the AI Act's requirements for risk-aware human oversight, transparency of decision-making, and prevention of over-reliance on automated outputs. In this way, the ChatMED-HITL platform provides a practical, reusable mechanism for embedding human oversight into the lifecycle of generative AI systems in healthcare.

1.2.3 *Work Package 3: Upgrading a Research Management/Administration Unit*

WP3 focuses on strengthening the institutional research management, administrative, and technical infrastructure at the Widening coordinator (FCSE), with the objective of enabling sustainable, state-of-the-art research activities in generative AI for healthcare. The work package addresses both **organizational capacity** (research support structures, governance, and knowledge management) and **technical readiness** (computational infrastructure), ensuring that the institution is prepared to carry out and sustain the research and innovation activities foreseen in the project and beyond.

T3.1 Benchmarking infrastructure against the state of the art. Task T3.1 was completed through the development of Deliverable D3.1 – Report on the upgraded infrastructure and knowledge management repository establishment. The task benchmarked FCSE/UKIM's existing research management and infrastructure against state-of-the-art practices at leading international institutions, namely Graz University of Technology (TU Graz), Technical University of Munich (TUM), Swiss Federal Institute of Technology Zurich (ETH Zurich), and Massachusetts Institute of Technology (MIT). The methodology combined (i) a comparative analysis of organizational structures, support units, governance mechanisms, and research performance indicators at the reference institutions, and (ii) a structured questionnaire assessing the current state at FCSE/UKIM. The insights gained from the comparative analysis led to the tailored questionnaire focusing on governance, support services, incentives, the talent pipeline, and research performance. These observations further informed a set of forward-looking recommendations aimed at enhancing the scalability of AI-oriented computational infrastructure, improving project-level research support capacity, and complementing existing institutional systems with domain-specific data and knowledge management workflows. These recommendations directly guided the strategic planning, infrastructure procurement, and knowledge management activities implemented in Tasks T3.2–T3.4. Ultimately, the broader university-level recommendations are designed as a strategic foundation for a phased, long-term implementation, ensuring the ChatMED project fosters a deep and lasting institutional modernization.

T3.2 Developing a strategic plan for infrastructure upgrade. Building on the findings of T3.1, FCSE developed a strategic plan for upgrading its computational and research infrastructure. The plan includes the procurement of an H100-based GPU server, which will be dedicated to the ChatMED project and integrated with the existing 8×A100 GPU infrastructure, significantly increasing the capacity for large-scale LLM training, evaluation, and fine-tuning. In parallel, the FCSE team initiated preparations for participation in a EuroHPC Joint Undertaking call to **establish an AI Antenna**, envisaged in connection with the Greek AI Factory. This strategic initiative aims to position FCSE/UKIM within the broader European high-performance computing and AI ecosystem, ensuring long-term sustainability, access to cutting-edge resources, and alignment with European AI infrastructure initiatives beyond the lifetime of the project.

T3.3 Procuring IT infrastructure (get ready for research). Task T3.3 focused on the practical implementation of the infrastructure upgrade plan. The procurement and deployment of the planned IT infrastructure were completed in early Q3, and the systems were fully operational and ready to support the research component of the project. This milestone ensured that FCSE had the necessary computational capacity, security, and reliability to execute the planned benchmarking, training, and evaluation activities involving large language models.

T3.4 Establishing a knowledge management system. Task T3.4 addressed knowledge management and long-term preservation of project outputs along two complementary directions,

both documented. The first direction focused on establishing a public, long-term preservation mechanism through the project's Knowledge Repository, integrated into the ChatMED website. This repository provides open access to project-generated materials, including reports, presentations, datasets, and other public outputs, supporting transparency, open science, and reuse.
The second direction extended beyond static archiving toward advanced scientific data management through the TriZ-Flow architecture. TriZ-Flow introduces a structured, lifecycle-oriented data management approach, encompassing data ingestion, transformation (including anonymization and standardization), and publication of FAIR-aligned datasets. An initial interface has already been developed, as depicted in Figures 3, and the system will be further refined with additional functionalities throughout the remainder of the project. Together, these mechanisms ensure that the data produced or gathered within ChatMED and any project in the future are not only preserved but also curated, governed, and made reusable for other research initiatives.

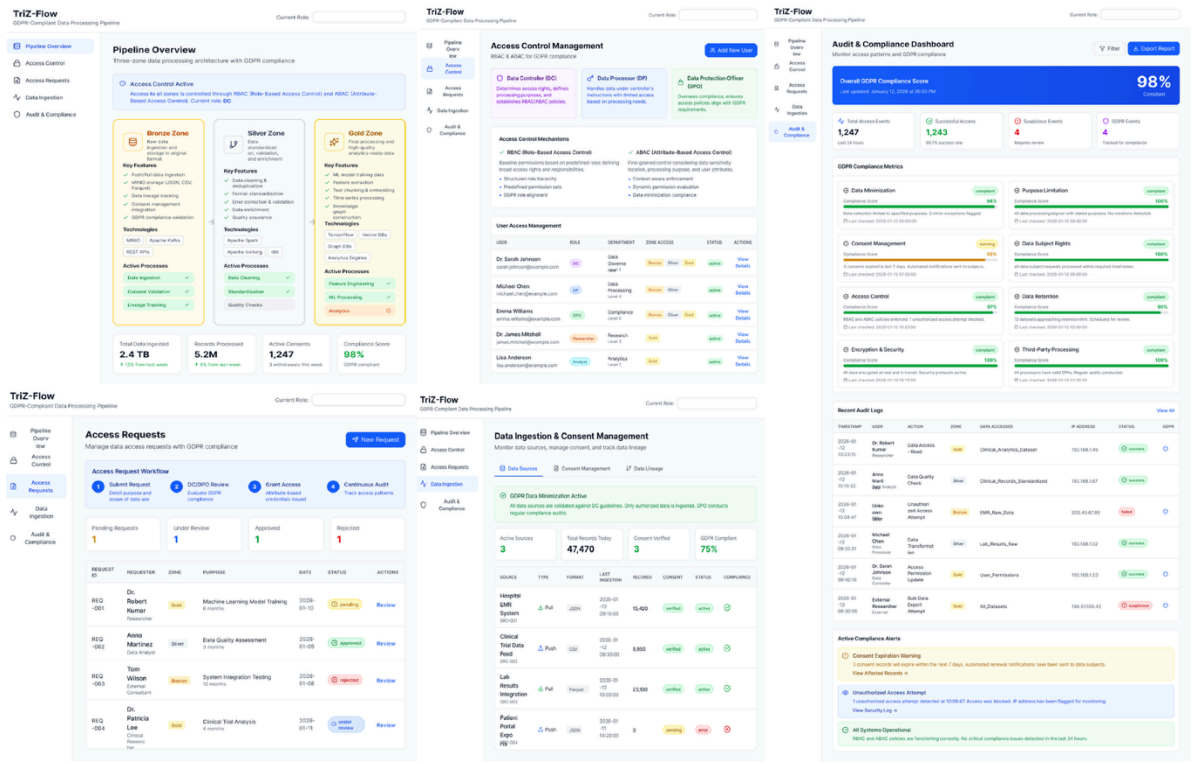


Figure 3. TriZ-Flow pipeline overview.

1.2.4 *Work Package 4: Strengthening Fundamental Research Skills*

WP4 focused on strengthening the fundamental research capacities of the consortium through structured training activities, targeted short-term staff exchanges (STSEs), an interdisciplinary summer school, and continuous CNDE actions. The work package was implemented in accordance with the plans defined in Deliverable D2.4, with all activities executed, documented, and consolidated in Deliverable D4.1. Collectively, WP4 activities enhanced technical, clinical, regulatory, and interdisciplinary competencies while directly contributing to downstream scientific, technical, and policy-oriented outcomes of the project.

T4.1 Executing fundamental training activities. This task implemented the structured training program defined in Deliverable D2.4, with execution and outcomes comprehensively reported in Deliverable D4.1. During the first project year, two major training sessions were organized by the coordinating institution (FCSE), combining theoretical foundations with hands-on workshops and industry engagement.

The training sessions covered the topics in Generative AI, including LLM fundamentals, prompt engineering, retrieval-augmented generation (RAG), domain-driven software design, multimodal AI, medical imaging, and AI applications in neurology. With those sessions we explicitly closed the training topics on: Generative AI Basics, Networking & Collaboration, Domain-Driven Software

Design, and AI & Neurological Disorders; as well as the hands-on topics: Hands-on Generative AI, Infrastructure & HPC, User Experience in Medical Applications, and Medical Imaging & AI.

The activities were deliberately designed to bridge gaps between AI researchers, clinicians, and industry stakeholders, ensuring practical relevance and immediate applicability. High participation levels, strong interdisciplinary engagement, and positive feedback confirmed that the training successfully strengthened fundamental research skills and fostered collaboration across domains, directly supporting the overall objectives of the ChatMED project.

T4.2 Executing STSE programs on solid bases for research. This task executed the STSE program as planned in Deliverable D2.4, with outcomes detailed in Deliverable D4.1. All planned STSEs were successfully implemented (Neurological Case Studies, Advanced Medical Imaging Techniques, and Innovative Algorithm Design) and one additional STSE originally foreseen for Year 2 (Medical Software Development Life Cycle) was advanced and executed during the first reporting period in response to emerging project needs.

The STSEs made a substantial scientific and strategic contribution to the project. Their key outcome was the establishment of the neurology-driven orchestrator foundations, including workflow design, modular agent specification, dataset preparation, and integration strategies, which now serve as the basis for further refinement and validation in WP7. In addition, one STSE led to the development of a strategic and exhaustive evaluation framework for vision-enabled large language models in the neurology domain, addressing multimodal clinical reasoning, interpretability, and reliability aspects. This work resulted in a completed research paper that will be submitted to an eminent journal.

Beyond technical development, the STSEs also triggered high-impact scientific and policy-oriented outputs. One exchange directly resulted in a publication in the JMIR Tutorial section addressing the implications of AI-driven clinical decision-making on medical reports and the need for AI annotations to remain compliant with MyHealth@EU requirements. Another STSE led to the creation and publication of a benchmark dataset in Elsevier's Data in Brief, strengthening the project's open science and evaluation footprint.

Overall, the STSE program significantly accelerated knowledge transfer, interdisciplinary collaboration, and tangible research outputs, exceeding the originally planned scope for the first project year.

T4.3 Executing summer school on interdisciplinarity. The summer school 1 (Interdisciplinary Innovations) was organized in line with the program defined in Deliverable D2.4, with its execution and outcomes reported in Deliverable D4.1. While preserving its original networking and interdisciplinarity objectives, the summer school organized by JSI evolved into an intensive co-creation environment focused on joint idea development and proposal preparation. Through structured brainstorming and thematic workstreams, the summer school facilitated deep collaboration between AI researchers, clinicians, regulatory experts, and industry partners. A major outcome of this activity was the initiation and preparation of two large follow-up project proposals as extensions of ChatMED, leveraging the scientific, technical, and organizational foundations established within the consortium. Further details on these extensions and their expected impact are provided in the Impact section of the report.

T4.4 Executing CNDE activities. CNDE activities under WP4 were carried out in accordance with the plan defined in Deliverable D2.2 and extended beyond the originally foreseen communication and dissemination actions to include broad networking and exploitation activities. These actions spanned domestic and international conferences, targeted workshops, policy-oriented events, journal publications, and sustained engagement with academic, clinical, industrial, and public-sector stakeholders. The outcomes of these activities are reported in Deliverable D4.1 and further consolidated in the Impact section.

A number of flagship activities illustrate the scope and reach of the CNDE actions. FCSE organised a dedicated ChatMED brainstorming workshop during the 16th ICT Innovations Conference (September 2024), bringing together young researchers and senior scientists to explore the application of large language models in medical diagnosis and treatment support. In addition, the "ChatMED in Medicine" session at the Information Society 2024 Multiconference in October 2024 in Ljubljana provided a multidisciplinary forum for clinicians, AI researchers, and engineers to

discuss practical integration of LLMs into healthcare systems.

A major Message Transfer Activity was held in November 2024, hosted by FCSE, with participation from high-level institutional and public stakeholders, including representatives of the Ministry of Health and university leadership. This event fostered dialogue across academia, healthcare, and industry, addressing robustness, explainability, trustworthiness, and ethical deployment of AI in medicine, and positioned ChatMED as a human-centric and ethics-driven initiative.

A distinctive highlight of the first project year was the thematic focus on cognitive wellbeing, expressed through an art–science exhibition comprising 16 visual artworks generated using carefully designed human prompts and ChatGPT-4o. Initially presented during the Message Transfer Activity, the exhibition has since been permanently installed across multiple university locations, reinforcing the project's long-term visibility and institutional embedding. Together, these CNDE activities significantly amplified the project's outreach, visibility, and societal impact, with full documentation available through D4.1 and the project website.

Figure 4. Insight into the first art exhibition with the aim to promote the project's objectives.

1.2.5 *Work Package 5: Strengthening Technical/Administration Skills*

WP5 focuses on strengthening the technical, administrative, and managerial capacities required to support complex, interdisciplinary, EU-funded research projects. The work package addresses the skills needed for effective project coordination, regulatory compliance, quality assurance, and sustainable technical operations. WP5 spans the period from **1 September 2025 to 31 August 2026**; therefore, at the time of this report, activities are **partially implemented**, covering the period up to **31 December 2025**.

T5.1 Executing technical/administration training activities. Task T5.1 was initiated as planned, with the first training session successfully completed during the reporting period. In October 2025, the ChatMED consortium convened at the Jožef Stefan Institute (JSI), Ljubljana, where a dedicated training workshop was organised in conjunction with the 28th Information Society Multiconference. The training targeted technical and administrative staff as well as senior researchers, with the objective of strengthening core competencies required for managing EU-funded research projects. The workshop addressed interdisciplinary project management, financial control in Horizon Europe projects, quality assurance procedures, and software engineering practices relevant to research infrastructures and AI systems. By combining administrative and technical perspectives, the session contributed to a shared understanding of project governance and operational requirements, laying the foundation for the remaining WP5 activities to be executed in the second project year.

T5.2 Executing STSE programs on innovative design. Task T5.2 has been partially executed with one exchange completed in May 2025, while the other remained as planned for May 2026, in line with the original project schedule. Administrative staff are planned to undertake a stay at JSI, aiming

to deepen expertise in applying EU regulations, particularly in the area of scientific data management, governance, and compliance. The execution and outcomes of this task will be reported in the Deliverable D5.1.

T5.3 Executing summer school on neurology and AI fusion. Task T5.3 is scheduled for May/June 2026 and has not yet been executed during the current reporting period. The planned summer school will focus on the fusion of neurology and artificial intelligence, addressing both scientific and operational perspectives. Its implementation and outcomes will be documented in D5.1.

T5.4 Executing CNDE activities. Despite the partial implementation status of WP5, a number of CNDE activities were already carried out during the reporting period, contributing to the project's visibility and strategic positioning.

ChatMED participated as a panelist at the Twinning Info Day organised within the framework of the Twinning project "Twinning for Excellence in Adaptive Edge AI (AIDA4Edge)" at the Faculty of Electronic Engineering in Niš on 15 September 2025, contributing to discussions on capacity building and excellence in EU-funded research.

The ChatMED consortium also made a strong scientific contribution at the 28th Information Society Multiconference (IS2025), held at JSI, where five peer-reviewed papers were presented and made publicly available through the project's Knowledge Repository. These contributions addressed advanced evaluation methodologies, privacy-aware AI deployment, and trustworthiness of AI systems in healthcare, reinforcing ChatMED's position beyond generic chatbot applications.

In addition, ChatMED participated as a panelist at the 4th AmCham North Macedonia Healthcare Conference in November 2025, contributing to discussions on "Recommendations for Improvement in the Year Ahead" and engaging with stakeholders from healthcare, industry, and policy domains.

Finally, on 4 December 2025, the ChatMED team participated in a strategic networking meeting in Paris, hosted by AP-HP (Assistance Publique – Hôpitaux de Paris) and Inserm (French National Institute of Health and Medical Research). This meeting served as a high-level networking and ideation platform, exploring future collaborations and laying the groundwork for potential COST Actions and Horizon Europe proposals aligned with the long-term objectives of the project.

1.2.6 *Work Package 6: Strengthening Advanced Skills*

Work Package 6 has not yet been initiated, as its implementation is scheduled for a later project phase (starting in September 2026). Preparatory planning has been completed under earlier work packages, and execution will commence as foreseen in the Grant Agreement.

1.2.7 *Work Package 7: Research & Innovation (Data)*

WP7 focuses on the development of novel, neurology-driven data methodologies that enable structured, explainable, and verifiable use of LLMs in clinical reasoning. While WP7 formally started in September 2025, preparatory research activities were initiated during the STSEs in May 2025. These early activities ensured continuity between capacity-building actions (WP4) and downstream research and innovation objectives, accelerating the transition from training to research implementation. The work in WP7 directly operationalizes the project's ambition to move from unstructured textual medical knowledge toward **reasoning-aware, machine-interpretable neurology datasets**, while preserving clinical validity, transparency, and human oversight.

T7.1 Neurology-driven scenarios creation. This task resulted in the design and validation of a novel multi-agent, reasoning-aware methodology for transforming neurology cases from narrative descriptions into structured, stepwise clinical reasoning workflows. The methodology, to be fully reported in Deliverable D7.1 is grounded in expert knowledge, established neurology books, handbooks, and case-based peer-reviewed teaching materials. Instead of treating neurology cases as static "textual correlations," the proposed approach explicitly models expert neurological reasoning by decomposing decision-making into modular stages, implemented through the NeuroOrchestrator pipeline. This pipeline transforms the case studies into complete reasoning traces by systematically

filling clinical gaps using structured logic, hypothesis generation, validation, and safety checks.

The NeuroOrchestrator consists of a sequence of well-defined agentic steps, including case extraction, emergency risk assessment, hypothesis generation and ranking, question generation and resolution, diagnostic test planning, treatment planning with contraindication checks, quality validation, and structured report export. Each step produces machine-readable JSON outputs, enabling transparency, traceability, and downstream validation. To validate methodological robustness, representative scenarios spanning **basic (educational), intermediate,** and **advanced (journal-level)** neurology use cases were processed through the full pipeline. Validation was supported by a dedicated interface enabling step-by-step inspection and human verification of intermediate reasoning outputs, using representative examples per scenario family, covering epilepsy, cerebrovascular emergencies, general neurology, and complex multisystem disorders. T7.1 will culminate in Deliverable **D7.1**, which formally establishes the NeuroOrchestrator methodology as the conceptual backbone for subsequent dataset construction, fine-tuning, and evaluation activities in the WP8.

T7.2 Scientific data management. Building on the validated methodology from T7.1, T7.2 addresses the systematic scientific data management of neurology-driven reasoning content. The task focuses on transforming heterogeneous neurology knowledge sources, into a single, coherent reasoning dataset. Using the NeuroOrchestrator framework, each source is processed to extract structured representations of clinical reasoning. The resulting outputs are encoded in standardized JSON schemas defined in Deliverable D7.1, ensuring consistency, interoperability, and readiness for reuse across downstream tasks such as fine-tuning, benchmarking, and human-in-the-loop evaluation.
The final outcome of T7.2 will be an openly available reasoning dataset (D7.2), divided into:

- Basic scenarios, targeting education-oriented and training use cases;
- Intermediate scenarios, and
- Advanced scenarios, derived from complex, journal-level neurology cases requiring higher-order clinical reasoning.

This dataset will be published through the project's repositories in line with FAIR principles and will directly support Deliverable D7.2, enabling transparent, verifiable, and clinically grounded research on neurology-focused LLMs.

1.2.8 *Work Package 8: Research & Innovation (GenAI)*

WP8 focuses on the systematic evaluation, selection, and optimization of generative and multimodal AI models for neurological applications. To effectively handle the complexity of neurological decision-making, multimodality will be addressed by proposing a multi-agent concept that targets the different domains which construct the decision in Neurology (presented in Figure 5). This architecture integrates multiple specialized systems, each focusing on a specific data modality or clinical domain (to be further researched in WP8), orchestrated by a central NeuroOrchestrator (which methodology was completed in D7.1/D7.2) that synthesizes insights from all agents to generate comprehensive diagnostic reports and treatment recommendations.

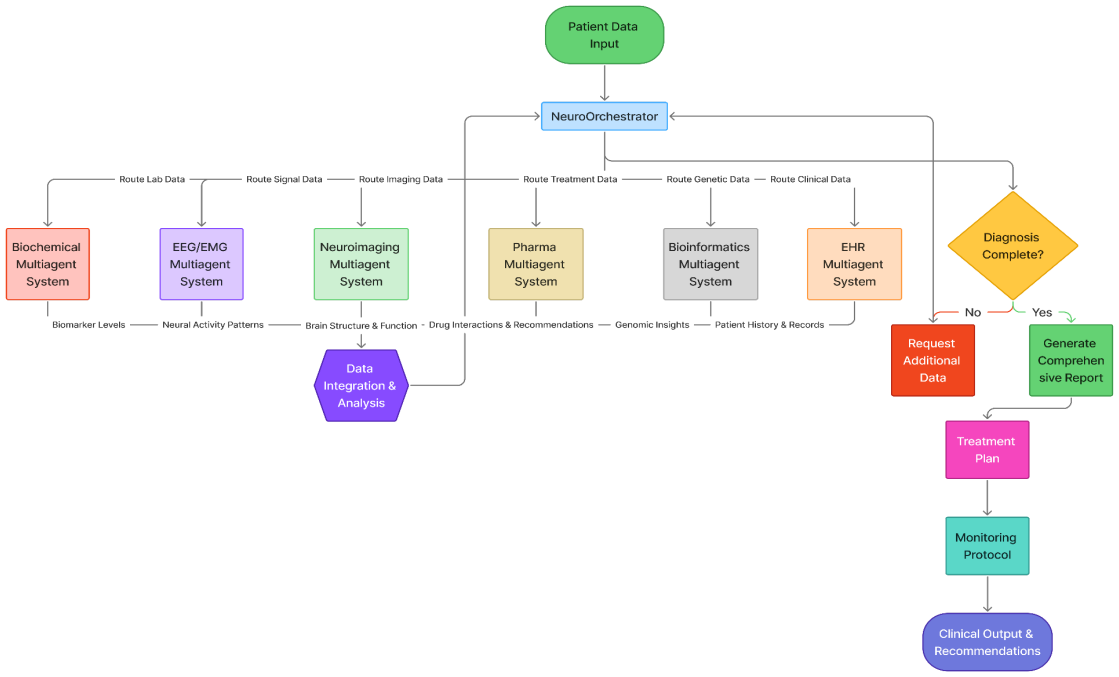


Figure 5. Addressing Multimodality through Multi-Agent Architecture

WP8 activities were initiated in December 2025 following preparatory work carried out during STSEs, enabling early validation of vision-enabled LLMs on fully labeled, publicly available neuroimaging datasets. Importantly, this validation phase is independent of WP7, as it does not rely on clinician-in-the-loop reasoning or newly generated scenarios, but instead leverages existing ground-truth annotations to establish a robust baseline for imaging model capability, reliability, and scalability. Also, some of the downstream tasks planned within WP2 have already been initiated and will serve as the basis for selecting the most suitable candidate models for fine-tuning, in line with the quality criteria defined in WP2.

T8.1 Data preprocessing. Work under T8.1 focused on the assembly, harmonization, and stratified partitioning of large-scale neuroimaging datasets. Multiple publicly available MRI and CT datasets were integrated from various sources (such as Br35H, AISD, Multiple Sclerosis MRI Dataset, and others), ensuring explicit diagnostic ground truth and expert-validated labels. Data were standardized across modalities, imaging sequences, and anatomical planes, and partitioned into support, development, and final test splits using stratified sampling to prevent data leakage and preserve class distributions. This preprocessing pipeline ensured reproducibility and enabled fair, controlled benchmarking across models and prompt regimes. The same approach will be used also for the other modalities.

T8.2 LLMs fine-tuning. As this task is scheduled from December 2025 until August 2026, we are at the very beginning where T8.2 activities focused on model screening and prompt-based adaptation. A unified prompting protocol was designed to enforce strict structured JSON outputs, mirroring real radiology reporting constraints. Models were evaluated under both zero-shot and few-shot regimes, with few-shot prompting using a fixed number of labeled examples per class to assess in-context learning potential. Results from this task directly inform the selection of candidate vision-enabled LLMs that will undergo targeted fine-tuning on narrower neurological tasks in subsequent project phases.

T8.3 Performance benchmarking. T8.3 represents the core completed activity of WP8 to date. A comprehensive multimodal benchmark was conducted on 20+ frontier and advanced vision-enabled LLMs, including proprietary and open-source models. Evaluation covered clinically meaningful tasks such as diagnostic classification, subtype identification, modality and sequence awareness, and anatomical plane recognition. Beyond accuracy and F1 metrics, the benchmarking framework explicitly assessed calibration, abstention behavior, structured output validity, computational

efficiency, and economic cost, providing a multi-dimensional view of model suitability for clinical and research deployment. The results demonstrated clear performance trade-offs, identifying a small subset of models that balance accuracy, robustness, calibration, and cost, thereby establishing a validated foundation for downstream optimization.

During the current reporting period, a benchmark dataset comprising 234 patient-oriented questions was established and used to conduct a comparative evaluation of 13+ large language models, focusing on their ability to support patient education through the delivery of clear, accurate, and accessible health information. This evaluation provides a systematic assessment of model performance in patient-facing communication and establishes a reproducible baseline for future analyses. In the upcoming period, the work will be extended to downstream tasks, including the evaluation of technical reliability with emphasis on system robustness and error handling, assessment of clinical history summarization capabilities, testing of medical knowledge inquiry and reasoning, exploration of prediction tasks related to clinical outcomes and disease progression in controlled research settings, and analysis of administrative support functionalities aimed at improving clinical documentation and workflow efficiency.

T8.4 Executing CNDE activities. Dissemination and exploitation activities under T8.4 have been actively pursued alongside technical work. The benchmarking methodology and results have been consolidated into a journal manuscript submitted to a high-impact venue, and intermediate findings have already informed consortium-wide decisions on model selection and infrastructure planning.

Overall work packages progress.

The following Table 7 presents an overview of activities implemented across the project work packages, together with the corresponding completion status of each WP, and association with the particular specific objectives.

Table 7. Work package implementation and completion status.

Work Package (WP)	**Status**	**Activities**	**Specific Objectives (SO)**
WP1	In progress	1. Establishment of project website and communication social and open science channels to keep all the activities (D2.3). 2. Project logo submitted for trademarking in December 2024.	All
WP2	Completed	1. DMP plan with an architecture for scientific data management platform established (D2.1). 2. Detailed communication, dissemination and exploitation plan established with tables to track the annual progress (D2.2). 3. Detailed training curriculum, STSEs and summer school programs are established with the focus on AI in neurology and mastering the project management (D2.4). 4. ChatMED HITL (D2.5) platform established and available as an open source code at ChatMED's Github.	SO2, SO4
WP3	Completed	1. Equipment with H100 GPU procured at the coordinator's facility. 2. AI Antenna established in connection with the Greek Pharos AI Factory.	SO1

		3. Benchmarking of the coordinator's organizational structure with the leading universities (D3.1). 4. Establishment of scientific data management pipeline (TriZ-Flow).	
WP4	Completed	1. Organized brainstorming session in September 2024. 2. Organized a special workshop at IS2024 in October 2024. 3. Organized message transfer activity in November 2024. 4. Organized art exhibition in November 2024. 5. Organized first training session in April 2025. 6. Organized industry panel in April 2025. 7. Launched STSE activities in May 2025. 8. Organized first summer school in June 2025. 9. Organized second training session in July 2025. 10. Submitted bilateral project between North Macedonia – Germany in June 2025. 11. Published 5 impact factors journals. 12. Published 10 conference papers.	SO2, SO3, SO4
WP5	In progress	1. Signed memorandums for collaborations with 4 industry partners in September 2025. 2. Organized third training session in October 2025. 3. Organized a special workshop at IS 2025 in October 2025. 4. Initiated and submitted new project proposal upon ChatMED results with 13 new partners at HORIZON-HLTH-2025-01-CARE-01 in September 2025. 5. Initiated and submitted new project proposal upon ChatMED results with 7 new partners at HORIZON-EIC-2025-PATHFINDERCHALLENGES-01-02 in October 2025. 6. Participated in a project proposal submitted to HORIZON-CL4-2025-03 in October 2025. 7. Participated in a COST action submission for the call OC-2025-1-28842 in October 2025. 8. Submitted bilateral project between North Macedonia – Austria in November 2025. 9. Participated as a panelist at the 4th AmCham North Macedonia Healthcare Conference in November 2025. 10. Participated in a networking event titled Future Healthcare Innovation at APHP & Inserm in Paris in December 2025. 11. Published 3 impact factors journals.	SO2, SO3, SO4

		12. Published 7 conference papers.	
WP6	Not yet initiated.	N/A	
WP7	In progress	1. Created methodology for mapping neurology case studies into reasoning-aware, machine-interpretable neurology datasets (to be delivered as D7.1). 2. Transformed heterogeneous neurology knowledge sources into a single, coherent reasoning dataset (to be delivered as D7.2).	SO5
WP8	In progress	1. Performed large-scale study on "Evaluation of Vision-Enabled Large Language Models for Clinical Reasoning in Neurological Disorders", currently submitted as a journal paper.	SO5

1.2.9 *Work Package 9: Ethics Oversight*

WP9 was introduced following the evaluators' recommendation to strengthen independent ethical oversight within the ChatMED project, in particular in relation to activities involving human participation and expert evaluation under T2.5 (Human Evaluation Framework). Given the project's focus on Large Language Models for clinical decision support, including interaction with vulnerable populations and the assessment of AI-generated medical content, dedicated ethics governance was deemed essential to ensure compliance with European and international ethical and regulatory standards throughout the project lifecycle.

WP9 therefore provides only deliverables providing independent ethics supervision, ensuring that the design, development, evaluation, and future deployment pathways of ChatMED remain aligned with GDPR, GCP, the Declaration of Helsinki, HIPAA (where applicable), and the emerging requirements of the EU AI Act.

D9.1 – POPD - AI - H - EPQ - NEC - Requirement No. 1. As part of WP9, the consortium established a formal ethics governance structure, including an internal ethics committee and the appointment of an independent External Ethics Advisor. The Ethics Advisor was specifically tasked with overseeing activities involving:

- human expert participation in model evaluation (T2.5),
- the use of medical data for training and benchmarking,
- interaction between LLMs and vulnerable individuals,
- and the ethical implications of AI-generated medical recommendations.

This task was completed within the first reporting period, ensuring that ethical oversight was in place before and during the execution of T2.5, rather than retrospectively.

D9.2 – AI - NEC - POPD - H - EPQ - Requirement No. 2. Deliverable D9.2 provides a comprehensive mapping between:

- project activities (notably D2.1 Data Management Plan and D2.5 Human Evaluation Framework),
- and the requirements of GDPR, ICH GCP E6, the Declaration of Helsinki, HIPAA, and the draft EU AI Act.

Key contributions of D9.2 include:

- validation of the privacy-by-design architecture (TriZ-Flow) and consent mechanisms,
- assessment of human oversight mechanisms embedded in the Human Evaluation Framework,
- identification and mitigation of algorithmic bias risks,
- and confirmation that no clinical decision-making is automated without human supervision.

Special attention is given to human-in-the-loop evaluation, confirming that expert judgment remains central at all stages of model assessment and that AI outputs are treated strictly as decision-support tools. The deliverable confirms that ChatMED is ethically prepared to progress toward future clinical validation phases while maintaining patient protection and regulatory compliance.

D9.3 – NEC - AI - H - POPD - EPQ - Requirement No. 3. D9.3 covers continuous ethics monitoring throughout the remainder of the project, culminating in a final ethics report at M36. It will:

- document all ethical issues encountered during the project,
- describe corrective or preventive measures taken,
- assess the ethical robustness of downstream research and innovation activities (WP7 and WP8),
- and provide a consolidated ethics summary supporting future clinical and exploitation pathways.

D9.3 will also reflect on the long-term implications of deploying generative AI in healthcare and provide recommendations for ethically responsible scaling beyond the project duration.

1.3 Impact

Strategic impact and TRL progression.

The ChatMED project has generated an impact that significantly exceeds its initial scope, positioning it as a strategic enabler for long-term research, innovation, and policy development in AI-enabled healthcare. As illustrated in Figure 5, ChatMED establishes a validated research foundation at TRL3–TRL4, while deliberately creating structured pathways toward higher TRLs through coordinated scientific output, infrastructure development, regulatory alignment, and sustained networking.

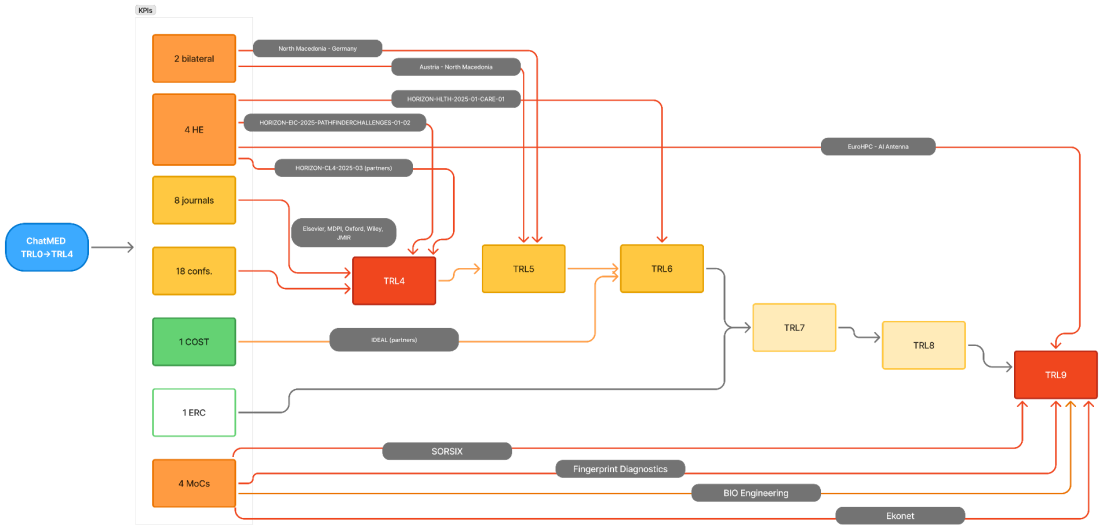


Figure 5. ChatMED TRL progress at mid-project.

Through a dense portfolio of peer-reviewed publications, benchmarking datasets, training actions, and multi-agent prototypes, the project has consolidated its scientific core and enabled downstream initiatives that naturally progress toward TRL5–TRL6 within follow-up Horizon Europe actions, bilateral projects, and industry collaborations. Importantly, this progression is not linear or isolated. ChatMED functions as a platform project, where multiple exploitation and continuation routes operate in parallel, reinforcing each other across scientific, technical, and organizational dimensions.

Scientific and technological impact.

From a scientific perspective, ChatMED has contributed novel methodologies and resources that are already shaping the research agenda in generative AI for healthcare. These include:

- systematic human-in-the-loop evaluation frameworks aligned with emerging AI Act requirements;
- benchmarking methodologies for LLMs, RAG pipelines, and vision-enabled models in neurology;
- modular multi-agent architectures that move beyond single-model experimentation toward coordinated reasoning systems; and
- openly accessible datasets and tools that enable reproducibility and reuse by the wider research community.

The project's outputs have directly enabled the formulation of new research questions and consortia, most notably the ChatMED-CARE and ChatMED-GLIA proposals, which extend the original neurology focus toward cross-domain clinical decision support and neuro-oncology. These proposals explicitly build on ChatMED's results, demonstrating a clear knowledge-to-proposal pipeline that strengthens Europe's capacity for high-risk, high-gain medical AI research.

Networking and ecosystem impact.

A defining element of ChatMED's impact is the breadth and depth of the network that has emerged around the project. Through sustained engagement with academic, clinical, industrial, and policy stakeholders, ChatMED has evolved into a recognised hub for discussions on trustworthy AI in healthcare.

At the national level, ChatMED has actively contributed to shaping strategic dialogue. The project's involvement as a panelist at the AmCham North Macedonia Healthcare Conference positioned ChatMED not only as a research initiative, but as a credible voice in discussions on the future of healthcare digitalisation, AI governance, and innovation-driven reform. This engagement strengthened links with decision-makers, industry representatives, and healthcare providers, reinforcing the project's relevance beyond academia.

At the European level, the invitation for ChatMED representatives to participate in a strategic networking meeting in Paris, hosted by leading medical research institutions, marked a significant recognition of the project's maturity. This event went beyond dissemination: ChatMED was explicitly invited to share experience, methodologies, and lessons learned, and to jointly explore new research directions with a broader European research community. As a result, new collaboration lines were opened, particularly in the context of future COST Actions and Horizon Europe proposals, extending ChatMED's influence to entirely new groups of researchers.

Industry engagement and exploitation pathways.

ChatMED has established strong and growing links with industry partners, ensuring that research outcomes remain grounded in real-world applicability. Collaboration with healthcare IT vendors, diagnostic companies, and infrastructure providers has enabled early alignment between scientific innovation and operational constraints, including data governance, interoperability, and regulatory compliance.

These industry interactions are not limited to technology transfer; they actively inform the project's research direction, ensuring that developed methodologies are deployable, scalable, and regulation-ready. The emergence of multiple follow-up proposals with explicit industrial participation demonstrates that ChatMED has succeeded in creating credible exploitation pathways that can support progression toward TRL7–TRL9 beyond the project's lifetime.

Long-term and systemic impact.

Taken together, ChatMED's impact lies not only in individual outputs, but in its role as a catalyst

project. It has:

- strengthened research capacity in a Widening country while embedding it in European networks;
- demonstrated how regulatory-aligned GenAI research can be conducted responsibly in healthcare;
- enabled sustained collaboration across academia, clinics, industry, and policy; and
- created a pipeline of follow-up actions that ensure continuity, scalability, and long-term sustainability.

By combining scientific excellence, structured TRL progression, active networking, and industry engagement, ChatMED contributes meaningfully to Europe's ambition to lead in trustworthy, human-centric AI for healthcare, while ensuring that its results continue to generate value well beyond the project's formal duration.

Table 8 presents a summary table that contrasts expected vs achieved KPIs.

Table 8. Impact KPIs: planned vs achieved (as of 31.12.2025)

KPI category	**Planned (whole duration)**	**Achieved (until 31.12.2025)**	**Evidence**
Training sessions	10 topics	6 completed	Training sessions 1 (April 2025), 2 (July 2025), and 3 (October 2025).
Workshops	10 topics	6 completed	Workshops within training sessions 1, 2, and 3.
STSEs	7 topics	4 completed	Five researchers were sent to JSI in May 2025 to work on the four topics.
Summer schools	3	1 completed	Completed summer school in June 2025 at JSI.
Peer-reviewed journal publications	10 at least	8 published	Published in high-quality venues including MDPI (Brain Sciences, Healthcare, Mathematics), Elsevier (Data in Brief), JMIR, Wiley (Proteomics), Oxford (Bioinformatics Advances), J Med Biochem.
Conference publications	30 at least	18 published	Papers at international conferences including Information Society 2024/2025, EEC 2024, CIIT 2025, ECNP 2025, BIOINFORMATICS 2026 (accepted, upcoming presentations in March 2026).
Benchmark datasets	Not explicitly quantified	1 published	RAGCare-QA dataset published; additional datasets under development (neurology reasoning, VLM benchmarking).

Horizon Europe proposals	3 at least	4 submitted, out of which 1 granted	ChatMED-CARE (HORIZON-HLTH-2025-01-CARE-01); ChatMED-GLIA (HORIZON-EIC-2025-PATHFINDERCHALLENGES-01-02); 3D-SMASH (HORIZON-CL4-2025-03); EuroHPC AI Antenna (EuroHPC JU).
Bilateral project proposals	3 at least	2 submitted	North Macedonia–Germany; North Macedonia–Austria
COST action proposal	1	1 submitted	IDEAL (OC-2025-1-28842)
ERC proposal	1	Not yet submitted.	Not yet initiated.
Industry collaborations / MoUs	5	4 signed	Sorsix, Ekonet, Fingerprint Diagnostics, Bio Engineering;
Art exhibitions	3	1 completed	Organized art exhibition on Cognitive Wellbeing during the Message Transfer Activity in November 2024.
Patent application	1	Not yet completed.	Not yet initiated. Trademarking initiated in December 2024.

A summary presentation reflecting the project's progress as of 31.12.2025 is available at: https://progress-report-1.chatmed-project.eu/.

2. Open Science

The ChatMED project follows a strong Open Science approach, ensuring early, open, and continuous sharing of research outputs throughout the project lifecycle. To this end, a dedicated ChatMED community on Zenodo[4] has been established and serves as the central archival point for all publicly shareable outputs, including peer-reviewed publications, pre-prints, work-in-progress manuscripts, public deliverables, benchmark datasets, and supporting materials. All items are versioned, citable, and assigned persistent identifiers (DOIs), ensuring long-term accessibility and traceability.

The consortium is fully committed to open access publishing. Scientific results are disseminated exclusively through open-access journals and conferences with open proceedings whenever possible. In parallel, the consortium systematically makes use of pre-print platforms, including arXiv, medRxiv, and bioRxiv, to enable early sharing of research findings, promote transparency during peer review, and accelerate knowledge transfer to the scientific and clinical communities.

To support open and reusable AI research, the project has established:

- A Hugging Face community[5], where benchmark datasets and trained or fine-tuned models are/will be published with accompanying metadata and usage documentation.

4 https://zenodo.org/communities/chatmed/records

5 https://huggingface.co/ChatMED-Project

- A GitHub repository[6], hosting all open-source software developed within the project. All repositories include version control, clear licensing, and README documentation.
- All publicly available outputs are additionally published through the project website Knowledge Repository[7], providing a single point of access for external stakeholders and ensuring coherence across dissemination channels.

3. DEVIATIONS FROM ANNEX 1 AND ANNEX 2 (IF APPLICABLE)

No formal deviations from Annex 1 or Annex 2 of the Grant Agreement have been implemented or approved during the reporting period.

3.1 Tasks/objectives

N/A.

3.2 Use of resources

NA for lump-sum grants

6 https://github.com/ChatMED

7 https://chatmed-project.eu/knowledge-repository/