DAIsy – Developing AI ecosystems improving diagnosis and care of mental diseases

ITEA 4 – 21016

Work package 4 (WP4) : AI Technology Development

Deliverable 4.4 : First versions and validation reports of collaborative AI models (doc & software)

Document type	: Deliverable
Document version	: No. 1
Document Preparation Date	: September 2025
Classification	: Confidential
Due Date	: October 2025

25 years ITEA 4 Community 1998–2023

Contents

I. Introduction

This deliverable provides the first consolidated status of the AI models developed within Task 4.4: Collaborative Intelligence: combining multiple AI technologies. The work is divided into four complementary subtasks aligned with the overarching goal of developing interoperable, explainable, and synergistic AI systems for mental health applications. For each subtask, the report summarizes objectives, partner contributions, current development status, validation approach, and key dependencies.

This document is restricted to collaborative AI in the context of DAIsy use-case and partner contributions.

II. Combining different data sources

Across the various subprojects, a central theme is the integration and preprocessing of heterogeneous data sources to enable advanced AI-driven analysis and decision support. Whether combining structural MRI with clinical findings, synchronizing EEG and fNIRS signals in real time, aggregating wearable and environmental data via mobile applications, or anonymizing sensitive clinical texts for privacy-compliant analytics, each initiative contributes to a broader ecosystem of multimodal data fusion. These efforts share a commitment to harmonizing diverse inputs—ranging from neuroimaging and biosignals to patient-reported outcomes and institutional records—into coherent, analyzable formats. This unified approach lays the foundation for robust model training, cross-platform interoperability, and clinically relevant validation strategies.

Partner-specific contribution

The **ARD Group** is developing a data integration and preprocessing pipeline for structural MRI and clinical findings. After gray matter segmentation using CAT12, the images are normalized to MNI space using DARTEL and smoothed with Gaussian filtering. 90×90 Kullback-Leibler similarity matrices are then derived using KDE. A second preparation pipeline for XTRACT data has been completed. These feature sets will be combined to train multiple AI models. Work is ongoing. Validation is based on clinician assessment and retrospective AI validation and depends on access to clinical data from the Hospital Information Management System (HIMS or HMS).

OFFIS is developing MultiPy, a multimodal real-time toolbox for electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) data preprocessing and analysis. The system enables real-time acquisition, synchronization, and preprocessing, including state-of-the-art artifact removal and filtering. It supports standardized communication via Lab Streaming Layer (LSL) and export to common neuroscience formats, including BIDS. The toolbox is under development. Validation is planned through simulation with pre-recorded datasets and subsequent studies. Dependencies include EEG/fNIRS hardware, LSL-based acquisition, and/or access to dataset repositories.

Ascora is developing a therapist dashboard that consolidates AI-based predictions of depression progression with data from the HMS patient app and other connected sources. The interface also supports scheduling activities outside of therapy sessions. The dashboard is a practical prototype currently under evaluation, validated through clinician feedback, and based on HMS and the patient app.

Ascora and **OFFIS** are collaborating on the patient app to support depression treatment to aggregate data from wearables, patient inputs, mobile devices, and environmental sources (e.g., weather by location). The app also offers learning content and tasks between sessions. It is a practical prototype currently under evaluation. Validation will be conducted through a feasibility study with volunteers and is dependent on the therapist dashboard from Ascora.

MEDrecord BV is implementing a privacy-compliant analytics pipeline using Google VaultGemma-1B (a differentially private LLM) for sensitive text data such as clinical notes, patient reports, and therapy transcripts. The pipeline applies automated anonymization of protected health data, ensures formal differential privacy guarantees (target: $\varepsilon \leq 2.0$, $\delta \leq 1.1e\text{-}10$), and generates privacy-compliant embeddings to enable cross-institutional collaboration. This project is planned or already underway. It will be validated through privacy checks (e.g., membership inference attacks), comparisons to non-private baselines, and regulatory compliance checks. It depends on access to clinical records by healthcare institutions.

III. Promoting AI-driven analytics innovations

The subprojects collectively advance AI-driven analytics by developing innovative, domain-specific architectures that leverage multimodal data and machine learning techniques to support mental health diagnostics and interventions. From ARD Group's classification models distinguishing bipolar from unipolar depression using structural MRI, to OFFIS's real-time neurofeedback module integrating EEG-fNIRS features for mental state classification, and MEDrecord BV's agent-based RAG system combining clinical narratives, physiological data, and behavioral patterns into a transparent decision-support framework—each initiative contributes to a growing ecosystem of intelligent, interpretable, and clinically relevant AI tools. These efforts emphasize modularity, transparency, and continuous learning, with validation strategies tailored to real-world clinical settings.

Partner-specific contribution

The **ARD Group** is currently developing a prototype MDD (Major Depressive Disorder)-BD (Bipolar Disorder) classification architecture that evaluates a range of classical and hybrid ML models using MRI-based features. Although the prototype phase is complete, the results are not yet generalizable; further work on feature fusion and validation is ongoing. The primary use case is differentiating between bipolar and unipolar depression using structural brain MRI.

OFFIS has developed a modular neurofeedback machine learning module within MultiPy that integrates different machine learning models such as support vector machines (SVMs), feature selection, and real-time general linear model (GLM) fitting to process fused EEG-fNIRS features for real-time applications. This functional prototype has initial benchmarks in simulation; empirical validation is pending. The system aims to classify mental states for neurofeedback/brain-computer interfaces (BCI) and works with real-time streams or BIDS-formatted historical recordings.

MEDrecord BV is developing MentalHealth-MERA and its TraceMind interface, an agent-based RAG (Retrieval-Augmented Generation) system with a citation-based continuous learning mechanism and a multimodal "expert council." Specialized models analyze the same patient case from different perspectives: a multimodal model for clinical images, a

clinical LLM for notes and narratives, a time series model for physiological patterns, and a behavior analyzer for app usage. Each AI output includes clickable citations linking to clinical cases, literature, or clinician feedback, allowing transparent review of reasoning paths. The agent retrieves similar historical cases and results, and a weighted consensus mechanism reconciles outputs and flags disagreements for expert validation. Clinician feedback is stored at the citation level, enabling continuous improvement without retraining. Internal validation shows a consensus accuracy of 0.88, an inter-model agreement of 0.76, a retrieval relevance of 0.82, and a clinical acceptance rate of 89%. The project is in progress and will be gradually deployed for decision support, focusing on transparency, traceability, and clinician trust. The use case is a comprehensive, interpretable assessment across all available modalities, with data coming from Hospital Information Management System (HIMS) and a vectorized knowledge base.

IV. Building synergistic AI networks (depending on II and III)

The subprojects are collectively laying the groundwork for synergistic AI networks by developing interoperable infrastructures that enable seamless coordination between diverse machine learning components. Whether through ARD Group's model repository supporting full ML lifecycles and REST-based deployment, OFFIS's modular architecture for real-time neurofeedback with interchangeable components, or MEDrecord BV's orchestration hub that dynamically triggers, aggregates, and reconciles AI outputs across modalities—each initiative contributes to a distributed, resilient, and scalable AI ecosystem. These efforts emphasize modularity, API-first integration, and robust validation strategies, ensuring that individual models and systems can operate collaboratively to support complex clinical workflows and decision-making.

Partner-specific contributions

The **ARD Group** is creating an AI (Artificial Intelligence) model repository that hosts models on MinIO and orchestrates end-to-end ML (Machine Learning) lifecycles: data acquisition, aggregation, cleansing, training, testing, and REST (Representational State Transfer) deployment. Work is in progress, oriented toward the WP5 backend and defining safe trigger output scenarios. Validation is operational and aims at complete data flows, robust cleansing rules, and successful pipeline executions. Dependencies include WP5 and data sources such as CSV/DICOM (Digital Imaging and Communications in Medicine) /MRI (Magnetic Resonance Imaging).

OFFIS has completed the modular neurofeedback system architecture, which supports real-time data flow integration via LSL. The individual modules (e.g., acquisition, preprocessing, classification, and feedback control) are extensible and interchangeable and feature Python/Matlab APIs for interoperability. The architecture is complete, but integration tests are pending. Planned validation includes latency, signal integrity, and module compatibility; dependencies include EEG/fNIRS hardware and compatible AI/control modules.

MEDrecord BV plans to establish a MentalHealth AI Orchestration Hub, a middleware service that coordinates multiple AI components across the consortium. It registers model features, triggers relevant models in parallel as new data arrives, aggregates results for consensus, and provides resilience (circuit breaker), versioning/rollback, and load

balancing. Target metrics include 99.9% availability, a cross-model agreement tracking rate of 0.78, and a throughput of 50 requests per minute. Integration is API-first and container-based (Docker/Kubernetes). The hub relies on MERA and partner interfaces.

V. Integrating human expertise with AI (depending on III or IV)

The subprojects share a common goal of embedding human expertise into AI systems to ensure transparency, usability, and clinical relevance. From ARD Group's clinician-facing dashboard for structured data entry and result visualization, to OFFIS's MultiPy GUI enabling real-time monitoring and manual control of neurofeedback experiments, and Ascora's therapist dashboard that combines AI predictions with practical planning tools—each initiative emphasizes human-in-the-loop design. MEDrecord BV's TraceMind interface further exemplifies this approach by allowing clinicians to inspect, correct, and contribute to AI-generated citations, fostering continuous learning without retraining. These systems are built not only to deliver intelligent outputs but also to empower professionals to guide, validate, and refine AI behavior through intuitive interfaces and feedback mechanisms.

Partner-specific contributions

The **ARD Group** is developing a clinical dashboard to support clinician data entry and result visualization. The interface is under development. Preparatory meetings for demonstrations and integration with real-world data are currently taking place. Validation will be based on feedback from clinicians and researchers, with results collected through user experience surveys.

OFFIS provides the MultiPy GUI for real-time monitoring of model results, adjusting thresholds, and flagging important events. The GUI enables transparent, human control of BCI and neurofeedback experiments. This is a practical prototype and is currently in the planning phase of being evaluated. Researcher-led sessions with real or simulated data are planned. Validation will be based on feedback from BCI researchers and task-based performance monitoring.

Ascora is further developing the therapist dashboard, which consolidates depression progression predictions with data from the HMS patient app and other AI outputs. The system supports administrative planning between sessions. It is a practical prototype in the evaluation phase, integrated into GNU Health as a demonstrator and using simulated data from the patient app and the HMS. Validation will focus on clinical-professional feedback, with insights gained from user surveys and a publicly accessible demonstrator.

MEDrecord BV is developing TraceMind, a citation-based continuous learning interface based on the MERA RAG platform. Each AI result contains inline citations pointing to concrete evidence—from retrieved clinical cases and literature to past clinician feedback and multimodal data points. Clinicians can analyze sources in detail, correct specific citations, and add interpretations, which the system immediately prioritizes for future queries (learning without retraining). The interface visualizes a "confidence gradient" (validated sources in green, disputed sources in yellow, new ones in gray) and detects citation conflicts requiring clinician mediation. Planned validation includes citation accuracy rates, speed of feedback

integration, clinician confidence scores (before/after), and longitudinal improvements in diagnostic accuracy. Initial testing indicates a 70% citation relevance rate using a multi-LLM as-judge process, with qualitative clinician feedback incorporated into the roadmap.

VI. Current Status and next steps

In Task 4.4, the partners are working on interoperable pipelines, robust AI modules, and user-centric interfaces. Data integration is progressing toward standardized, BIDS-compliant results and privacy-preserving text analytics. Model development focuses on explainability and reliability, with increasing use of consensus mechanisms and uncertainty measures. Network-level orchestration is being designed to ensure modularity, resilience, and traceability. Human-AI interfaces emphasize transparency and rapid feedback, enabling continuous improvements without disruptive training cycles.

The next steps include

1) extending simulated to empirical validations in MultiPy-based pipelines;
2) completing feature fusion for structural MRI-based classification;
3) tightening privacy audits for VaultGemma-based text analytics;
4) conducting orchestration pilot projects to test consensus workflows;
5) conducting clinician-in-the-loop evaluations of dashboards and TraceMind to quantify the impact on trust, speed, and decision quality.

This consolidated view will be updated as components are moved from prototype to validated versions and clinical partners report results from pilot implementations.