Clara

Center for Artificial Intelligence and Quantum Computing in System Brain Research

Deliverable 2.8

Title: Theme Development Meetings

Funded by the European Union

Project Number: 101136607

Project Acronym: CLARA

Call: HORIZON-WIDERA-2023-ACCESS-01

Project title	**Center for Artificial Intelligence and Quantum Computing in System Brain Research (CLARA)**
Grant Agreement Nr.	101136607
Funding scheme	HORIZON Coordination and Support Actions (CSA) Call: HORIZON-WIDERA-2023-ACCESS-01 Topic: HORIZON-WIDERA-2023-ACCESS-01-01-two-stage - Teaming for Excellence
Project duration	1 November 2024 - 31 October 2030 (72 months)
Project Coordinator	INDRC – International Neurodegenerative Disorders Research Center
Deliverable number	**D2.8**
Title of the deliverable	**Theme Development Meetings**
WP contributing to the deliverable	**WP2**
Deliverable type	**DEC** R: document, report; DEM: Demonstrator, pilot, prototype, plan designs; DEC: Websites, patents filings, press & media actions, videos, etc.; OTHER: Software, technical diagram, etc.
Dissemination level	**PUB** PU = Public, fully open, e.g. web; CONFIDENTIAL-C; RESTRICTED-R; SEN – Sensitive
Due submission date	**31 March 2026 (M17)**
Actual submission date	31 March 2026 (M17)
Author(s)	Laura Brody, Eva Dočkalová, Ara Khachaturian
Internal reviewers	Jean-Marie Bouteiller
Final approval	Václav Snášel

Disclaimer

Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Document Status – History of Changes

Version	Date	Contributor(s)	Description of Changes
v01	23/3/2026	Eva Dočkalová, Laura Brody	First draft created.
v2.0	30/3/2026		Approved at the WP level
v3.0	30/3/2026	Eva Dočkalová	Approved at the project level
final	31/3/2026	Václav Snášel	Approved by the CLARA Director

Document Keywords

AI Architectures; Large language models; Multimodal, longitudinal disease modelling; White paper

Confidentiality

Does this report contain confidential information?	Yes ☐ No ☑
Is the report restricted to a specific group?	Yes ☐ No ☑ *If yes, please precise the list of authorised recipients*:

Table of Contents

Executive Summary

The Theme Development Meetings are planned as annual workshops designed to identify current trends and needs in Alzheimer's disease research. These thematic, stakeholder-focused events—held virtually and/or in person—will be structured to foster emerging dialogue between the research community and policymakers (e.g., EU officials) across Europe and globally.
Deliverable D2.8 (Theme Development Meetings) aims to present the first of these meetings and to summarize the main outcomes of the scientific discussion on the topic *"AI Architectures for Longitudinal Disease Research: Validation, Governance, and Strategic Positioning"*.

1. Introduction

The first Theme Development Meeting was held on 6 March 2026 as part of the INDRC Scientific Advisory Board (SAB) meeting, with the participation of the following experts:

Attendees: The INDRC Advisory Board participants included the Chairman, Prof. Jean-Marie Charles Bouteiller, PhD, alongside Ara S. Khachaturian, PhD, and members Steven Carroll, MD, PhD; Jonathan Helfgott; Zaven Khachaturian, PhD; Louis Kirby, MD; Doc. Ing. Václav Křemen, PhD, MBA; Pei-Jung Lin, PhD; Michael W. Lutz, PhD; Nico Pronk, PhD; Josef Šivic, PhD; Prof. RNDr. Václav Snášel; and Peter J. Snyder, PhD. The Chairman of the INDRC Executive Board, James Arthur Rice, PhD, and Mgr. Priya Blažková-Chandra were also present, together with guest presenter Gully Burns, PhD.

Guest Presenter: Dr. Gully Burns is an agentic AI engineer specializing in solutions that leverage structured knowledge. He studied neuroscience at Oxford, where he developed systems for modeling neural connectivity and complex experimental design. In 2018, he transitioned to industry, working at the Chan Zuckerberg Initiative (now Biohub), and is currently exploring opportunities in the AI-for-Science domain.

2. Report on the Theme Development Meeting

Main topic: AI Architectures for Longitudinal Disease Research: Validation, Governance, and Strategic Positioning

2.1. Background and Objectives

On March 6th, 2026, the first in a planned series of Theme Development Meetings linked to the CLARA project took place as part of the INDRC Scientific Advisory Board (SAB) meeting. This event marked the launch of a broader initiative within the CLARA project aimed at developing workshops and expert discussion forums bringing together multiple stakeholders to identify current trends and needs in Alzheimer's disease (AD) research, and to align these with advances in artificial intelligence.

The topic proposed for discussion, "*AI Architectures for Longitudinal Disease Research: Validation, Governance, and Strategic Positioning*," was chosen due to the convergence of two key developments in biomedical research: (i) the rapid evolution of artificial intelligence (AI) architectures, particularly foundation models (e.g., large language models trained on large corpora), and (ii) the expansion of longitudinal and continuous health data, including electronic health records, claims data, imaging, molecular data, and increasingly, high-frequency biosensors and remote monitoring streams. These developments are enabling more granular data collection and are crucial for longitudinal analysis of disease progression.

Given CLARA's focus on multimodal, longitudinal disease modelling, there is a need to consider not only which AI architectures are most appropriate for such data, but also how these systems should be validated, governed, and strategically positioned. This line of inquiry was initially prompted by interest in applying AI methods to large-scale clinical datasets, which led to a broader examination of data structures and relevant modelling approaches. As part of this process, the CLARA team began developing an internal review of AI models spanning both foundation models and more traditional methods, such as survival analysis. While initial progress has been made in identifying suitable modelling approaches for multi-scale, multimodal longitudinal data, key questions remain regarding governance and sovereignty in the use of foundation models, as well as the strategic positioning of the INDRC and CLARA within the evolving AI ecosystem.

The purpose of this first Theme Development Meeting, organised around a guest presentation by Dr. Gully Burns and subsequent discussion, was therefore to contribute to an ongoing effort to gather input and refine these directions. An important question that emerged in preparatory discussions was whether to focus narrowly on large language models (LLMs) or more broadly on a range of AI architectures relevant to longitudinal disease research. While LLMs and Transformer-based foundation models are prominent in current AI research and discourse, they represent only one architectural family within a broader modelling landscape. For CLARA's purposes, given the involvement of multimodal, multi-scale, and potentially high-frequency disease data, a comparative architectural perspective – spanning classical longitudinal models as well as recurrent, attention-based, and state-space approaches – is more appropriate than an LLM-only focus.

Accordingly, the meeting's theme emphasized validation frameworks, governance, and strategic positioning rather than identifying a single optimal model class, with the aim of supporting an initial alignment of concepts, terminology, and strategic priorities across disciplines. This effort was intended not only to shape CLARA's internal direction, but also to contribute to a principled public framing of AI architectures, validation, and governance in longitudinal neuroscience research. As such, the discussion also aimed to inform the development of a public-facing opinion paper, which is expected to synthesize architectural considerations for longitudinal disease research, articulate validation and governance principles, and clarify strategic positioning for the INDRC and CLARA, with an emphasis on principled framing, interoperability, and responsible innovation.

2.2. Guest Presentation: Skillful Alhazen

The Theme Development Meeting opened with a guest presentation by Dr. Gully Burns titled "Skillful Alhazen: AI Agents for Scientific Knowledge Curation," providing a concrete starting point for the discussion. Dr. Burns, an AI researcher specializing in agentic systems and structured knowledge approaches in scientific domains, presented a preliminary implementation of an open-source, agent-based biocuration system designed to support interaction with scientific literature. His talk situated this work in relation to emerging 'deep research' AI tools, highlighting their usefulness but also their limitations as largely black-box systems. In response, he introduced 'Skillful Alhazen,' a system that aims to enable more transparent, structured, and traceable engagement with scientific knowledge through agent-driven workflows and structured representations of evidence.

Burns noted that tools such as ChatGPT enable users to gather and synthesize scientific information at scale, a capability that has rapidly proliferated across AI platforms and generally produces high-quality outputs. However, these systems remain largely opaque: while their results are useful, they offer limited transparency into the reasoning and evidence behind them, requiring users to critically evaluate the outputs, examine the evidence, and question the underlying assumptions. As a result, their output is effectively consumable rather than extensible. The Skillful Alhazen project, therefore, seeks to expose the process of 'deep research' and enable direct access to the underlying evidence, assumptions, and reasoning of AI-generated outputs.

Building on this, Burns proposed a shift from review-style outputs towards structured representations of knowledge. In this approach, entities, relationships, and sources are transparent, with their provenance visible and traceable. They are also explorable, allowing researchers to interrogate specific claims, identify gaps, and examine conflicting evidence. The system is designed to be extensible, supporting the integration of new domains, data sources, and reasoning processes without rebuilding. Finally, it is 'living', in the sense that the knowledge base grows and improves as more information is processed. Rather than simply generating answers, the aim is to construct a structured, evidence-rich knowledge representation that researchers can actively examine, question, and scrutinise. This reflects a broader argument in the talk: that scientific reasoning cannot be fully captured through natural language alone, but requires access to underlying structure, semantics, and reasoning processes. Instead of treating AI as a black box that produces answers, the goal is to expose and engage directly with its reasoning process and underlying representations.

This approach is contextualised by the notion of the 'bitter lesson' in AI: the idea that methods that scale with data and computational power tend to outperform those based on specialised, handcrafted models. In contrast, Burns suggests that frontier AI systems can be used to develop domain-specific understanding at scale. In this context, *Skillful Alhazen* uses Claude, or OpenClaw, as its underlying system and is accessed through a conversational interface on common messaging platforms. It operates through 'skills', which Claude describes as simple text instructions and Python scripts that are made available to an underlying agent. This agent can carry out actions such as coding, web search, and API use based on those instructions and its own intelligence and is recursive to some extent in that it can generate and refine its own skills. These skills also support domain modelling, enabling the system to construct and work with structured representations of knowledge.

A central component of the system is the use of a skill-specific TypeDB schema, a logic-oriented database that supports more expressive reasoning than conventional knowledge graph systems. While such databases typically require significant expertise to design and query, Burns emphasizes that this complexity can be mitigated by leveraging Claude's capabilities to generate and refine schemas and models. The system is designed as a personal curation environment that produces interactive dashboards through which users can access, organize, and explore information drawn from scientific literature and external data sources.

Building on this infrastructure, Burns emphasizes that the scaffolding and framing of the domain are key to how agentic systems operate. The system is therefore organized around a five-layer data model to represent domain knowledge. This comprises collections, which group related entities; entities, which represent the core domain objects; relations, which capture connections between entities; artifacts, which represent source information about those entities and preserve their provenance; and fragments, which are specific, relevant pieces drawn from those artifacts. These elements are complemented by notes, which record annotations and interpretations generated by either the model or a human curator. Together, these structures function as a scientific notebook or personal curation environment, in which information is recorded, annotated, and systematically organized.

This is complemented by a structured workflow for scientific curation, which serves as the system's task scaffolding. Curation is framed as a process consisting of five stages: foraging, in which relevant sources are identified; ingestion, where content is captured with associated provenance; sensemaking, where the model extracts entities, relationships, and insights from the material; analysis, which involves reasoning over the resulting representations to identify patterns and gaps; and reporting, where findings are synthesized and presented through dashboards and other outputs. This workflow maintains a clear distinction between source material and its interpretation, while leveraging the model's ability to organize and interpret complex information.

Underlying this workflow, Burns returned to the role of TypeDB as the system's core database, emphasizing its importance in enabling structured, transparent, and traceable knowledge representation. Unlike standard knowledge graph approaches, TypeDB supports richer representations of relationships, allowing additional metadata – such as provenance and confidence – to be attached directly to relationships. It also provides a flexible type system and supports composable reasoning, while maintaining a strongly defined schema that ensures data remains coherent and well-structured as it evolves. Together, these features enable more expressive reasoning over curated knowledge and reinforce the system's emphasis on transparency and reliability.

This framework ultimately gives rise to an agent-driven curation loop, in which the system operates as a self-improving cycle. Through interaction with the agent, the process iteratively develops, with each stage of curation contributing to subsequent analysis and interpretation. The system is designed to operate at scale, processing large volumes of literature, while supporting automated reasoning over the resulting knowledge structures. Notes play a central role in preserving understanding, providing human-readable context and interpretation, and identifying gaps that guide further investigation. In scientific use, this results in a structured, evidence-rich knowledge base anchored to standard ontologies, with transparent provenance, confidence measures, and the ability to query across domains to generate new hypotheses and insights.

Burns concluded the presentation with a reflection on the strengths, limitations, and future directions of the *Skillful Alhazen* approach. Its strength lies in integrating Claude with TypeDB, enabling rapid development of structured models, and in using lightweight 'skills' that support flexible, extensible workflows. The system also benefits from its iterative, self-improving design and its ability to incorporate externally defined tools. However, it remains in an early stage of development, is computationally expensive due to high token usage, and depends heavily on the model's domain understanding, highlighting the need for more robust evaluation. Future work will focus on developing domain-specific capabilities and strengthening engagement with the broader AI-for-science community to further refine and validate the approach.

2.3. Discussion: Key Themes and Implications

Following the presentation, the discussion among INDRC SAB members broadened to consider the implications of these approaches for AI architectures in longitudinal disease research, with particular focus on validation, governance, and strategic positioning. The discussion engaged with key questions around which models or combinations are most appropriate for multi-scale, multimodal longitudinal data, how to address governance and trust in the use of frontier AI systems, and how the INDRC and CLARA should position themselves within the evolving AI ecosystem. This included consideration of whether CLARA and the INDRC should act primarily as users of existing AI systems, contributors to model development, or as actors defining standards, validation frameworks, and infrastructural approaches. Throughout the discussion, emphasis remained on identifying a single optimal architecture rather than on the practical conditions under which different approaches could be used responsibly, particularly with respect to validation, data quality, and benchmarking. Participants drew on both practical experience and domain-specific challenges in Alzheimer's disease research, highlighting tensions around data heterogeneity, temporal scale, reliability, and interpretability, as well as the opportunities presented by increasingly powerful AI tools.

A key challenge concerned the application of AI approaches to existing longitudinal Alzheimer's disease data, which span multiple modalities but remain highly heterogeneous across patients. Participants highlighted the challenge of identifying combinations of early indicators that could reliably predict later patterns of cognitive decline, particularly given the need to link data across different time scales. This raised a broader issue around data integration, with emphasis on the difficulty of bringing complex, fragmented datasets into a coherent form suitable for analysis. It was noted that this represents a longstanding problem in knowledge representation and that recent advances in LLMs and agent-based systems may offer new ways to map diverse data sources into shared or federated schemas, enabling comparison and reasoning across datasets. At the same time, participants emphasised important constraints, including the limited availability of early-life data, the need for careful study design typically addressed through biostatistical expertise, and the importance of ensuring coherence between datasets. While AI systems were seen as potentially powerful in this context, particularly if configured to support structured analysis, it was stressed that human oversight remains essential.

Another key theme in the discussion was the practical use of AI tools in data analysis, highlighting both their potential and their limitations. Participants shared examples demonstrating that AI systems can produce sophisticated and highly useful outputs, including performing complex statistical analyses and identifying appropriate computational tools. However, these examples also revealed important risks: models may overlook fundamental aspects of a problem, leading to technically well-developed yet ultimately incorrect results. While the use of AI in multi-step analytical workflows can significantly reduce time and effort, it was noted that substantial manual oversight remains necessary, with researchers often needing to identify and correct both obvious and subtle errors. At the same time, participants highlighted the potential of agent-based workflows to improve reliability, with multiple agents performing complementary roles such as code generation, validation, and error checking. Overall, AI-driven approaches to data analysis were seen as a disruptive technology with significant potential for research productivity, but not yet fully reliable in practice, requiring systemic validation checkpoints throughout the workflow, potentially supported by agent-based systems to improve reliability.

A further theme concerned the expanding capabilities of AI agents in automating research workflows. Participants noted that recent advances enable agents not only to analyze data, but also to read scientific papers, implement described methods, train models, and evaluate results, potentially accelerating research tasks that would otherwise require substantial time and effort. However, this increased automation raises important concerns, including the persistence of errors and the risk that researchers may become more removed from the underlying data. This led to broader questions around trust and validation, with emphasis on the need to critically interrogate model outputs rather than accept them at face value. In this context, participants highlighted the growing importance of well-defined datasets, benchmarks, and problem formulations, which provide the basis for evaluating performance and guiding the effective use of these systems. It was also noted that data remains foundational to model development, and that measures such as confidence in inferred relationships may help identify gaps and guide further data collection or experimentation.

Finally, attention turned to emerging challenges related to high-dimensional and high-resolution data, particularly as increasing sensor availability is expected to generate larger and more complex datasets. Participants noted that addressing this shift will require adapting existing AI architectures, including agent-based approaches, to effectively model and integrate such data. At the same time, challenges related to data quality were highlighted, including issues of imbalance and incomplete or inconsistent labelling. It was suggested that AI and agentic systems may play a role not only in analysis but also in improving data quality, for example, through more efficient identification, labelling, and validation of difficult or underrepresented cases. Participants emphasized the importance of developing a public-facing opinion paper, seen as a timely and potentially impactful contribution in a rapidly evolving field where standards are not yet well established. Participants also noted that these developments extend beyond technical considerations, with implications for research practice, researcher training, and ethical frameworks in the use of AI.

2.4. Summary of Main Points

- The discussion highlighted the need for a comparative and integrative approach to AI architectures for multimodal, longitudinal disease data relevant to Alzheimer's disease research, rather than a focus on any single model class.
- The effective use of AI in research workflows was seen to depend on robust validation practices, given that current systems can produce technically sophisticated but potentially unreliable results without careful oversight, particularly in clinical and longitudinal research settings.
- Agent-based approaches were identified as a promising framework for structuring complex analytical workflows, particularly when combined with explicit validation and error-checking mechanisms.
- Progress in AI-driven longitudinal research was discussed as being constrained not only by model capability, but also by challenges in data integration, coherence, and availability, particularly in heterogeneous clinical datasets characteristic of Alzheimer's disease research.
- The use of foundation models raises important and unresolved questions around governance, transparency, and data sovereignty, which are central to the strategic positioning of the INDRC and CLARA within the evolving AI ecosystem.

- The discussion pointed to an opportunity for the INDRC and CLARA to contribute to the field by articulating frameworks for AI architectures, validation, and governance, including through a public-facing opinion paper and related activities.

Conclusion

The summary of the first Theme Development Meeting has been released on the CLARA website:
https://www.clara-center.eu/research-and-education/theme-development-meetings

In addition to this report, a white paper titled “A Scientific Operating System for Multimodal Longitudinal Disease Modeling” is currently in preparation and will also be made publicly available on the website.