



Funded by
the European Union



HORIZON-WIDERA-2023-ACCESS-02

ChatMED

Grant ID: 101159214

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



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



Jožef
Stefan
Institute



Univerzitet u Nišu
MEDICINSKI FAKULTET



Deliverable D7.1

Report on the Neurology-Driven Scenarios Created



Deliverable D7.1 - Report on the Neurology-Driven Scenarios Created			
Date:		January 31, 2026	
Lead Beneficiary:		FMN	
Type:		R	
Author(s):		Stevo Lukic, Ivana Vichentijevikj, Kostadin Mishev, Katarina Trojacanec Dineva, Monika Simjanoska Misheva	
Dissemination level:		PU	
Version:		1.0	
Abstract: This deliverable presents a multi-agent, reasoning-aware methodology for transforming narrative neurology cases into structured, stepwise clinical reasoning workflows. Grounded in expert neurological knowledge and established case-based literature, the approach explicitly models clinical reasoning by decomposing decision-making into modular, interpretable stages. These stages are implemented through the NeuroOrchestrator pipeline, comprising agentic steps for case extraction, risk assessment, hypothesis generation and ranking, question resolution, diagnostic planning, treatment planning with safety checks, quality validation, and structured report export. Each step produces machine-readable JSON outputs, enabling transparency, traceability, and reproducibility of the reasoning process. To ensure methodological reliability, the deliverable introduces a human-in-the-loop validation platform that verifies agent-specific prompt alignment and faithful mapping of free-text clinical narratives into structured representations. The resulting methodology establishes a controlled foundation for constructing and validating neurology-driven reasoning datasets (to be published in D7.2), which will be used to fine-tune the NeuroOrchestrator within the ChatMED multi-agent system (WP8).			
Disclaimer: The project is funded by the European Union under Horizon Europe Widening participation and spreading excellence programme, grant ID: 101159214. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.			
HISTORY OF CHANGES			
Date	Issue	Author	Summary of main changes
13.11.2025	1.0	Monika S. Misheva	Establishment of the template.
17.11.2025	2.0	Monika S. Misheva	Transfer of the NeuroOrchestrator methodology developed in May 2025 with examples.
15.1.2026	3.0	Ivana Vichentijevikj	Updating the methodology and providing new content.
30.1.2026	4.0	Stevo Lukic	Finalizing the methodology, check, and approval.



Table of Contents

List of Figures.....	4
List of Tables	4
List of Acronyms.....	4
Introduction	6
Methodological Approach to Framework Derivation	6
Clinical Neurology Workflow	7
The NeuroOrchestrator Agentic Workflow.....	9
Step A0: Case Extraction	9
Step A1: Emergency Risk Assessment.....	10
Step A2: Hypothesis Generation	10
Step A3: Question Generation.....	11
Step A3.1: AnsweredQAs	11
Step A4: Test Planner.....	11
Step A5: Treatment Planner.....	11
Step A5.5: Contraindication Checking (Optional).....	11
Step A6: Quality Validation.....	11
Step A7: Report Export.....	12
Hypothesis Refiner	12
Prompt-Based Agentic Mapping of Free-Text Clinical Narratives	12
The Neurology Scenarios Mapping	27
Neurology Scenario Example	27
Validation Platform for Agent-Specific Mapping Accuracy	40
Platform Overview.....	40
Case-Level Validation Workflow.....	41
Agent Output Inspection and Validation.....	42
Role of the Platform in the Methodology Refinement	44
Conclusion.....	45



List of Figures

Figure 1. Clinical neurology approach.....	7
Figure 2. Modified clinical neurology workflow.....	8
Figure 3. NeuroOchestrator workflow.	9
Figure 4. The case-centric dashboard.	41
Figure 5. Case-level view.....	42
Figure 6. Agent-level view.....	42
Figure 7. Agent validation example (a).	43
Figure 8. Agent validation example (b).	43
Figure 9. Agent validation example (c).....	44

List of Tables

Table 1. NeuroOrchestrator prompt engineering strategy.....	12
Table 2. Stepwise JSON schemas.....	23
Table 3. Neurology scenario example.....	28

List of Acronyms

AI (Artificial Intelligence)
 AAN (American Academy of Neurology)
 API (Application Programming Interface)
 CBC (Complete Blood Count)
 CDA (Clinical Document Architecture)
 CNDE (Communication, Networking, Dissemination, and Exploitation)
 COST (European Cooperation in Science and Technology)
 CrCl (Creatinine Clearance)
 CSF (Cerebrospinal Fluid)
 CT (Computed Tomography)
 EC (European Commission)
 EEG (Electroencephalography)
 EMG (Electromyography)
 EAN (European Academy of Neurology)
 ERC (European Research Council)
 EU (European Union)
 FAIR (Findable, Accessible, Interoperable, Reusable)
 FCSE (Faculty of Computer Science and Engineering)
 FHIR (Fast Healthcare Interoperability Resources)
 FMN (Faculty of Medicine Niš)
 FRAND (Fair, Reasonable and Non-Discriminatory)
 GA (Grant Agreement)



GDPR (General Data Protection Regulation)
GCS (Glasgow Coma Scale)
HEF (Human Evaluation Framework)
HITL (Human-in-the-Loop)
HL7 (Health Level Seven)
HTLV-1 (Human T-lymphotropic Virus Type 1)
ID (Identification)
ICP (Intracranial Pressure)
ICH (International Classification of Headache Disorders)
IV (Intravenous)
JSON (JavaScript Object Notation)
JSI (Jožef Stefan Institute)
LLM (Large Language Model)
LP (Lumbar Puncture)
LFTs (Liver Function Tests)
MRI (Magnetic Resonance Imaging)
NCS (Nerve Conduction Studies)
NICE (National Institute for Health and Care Excellence)
OMOP (Observational Medical Outcomes Partnership)
PMH (Past Medical History)
PNS (Peripheral Nervous System)
PO (Project Officer)
QA (Question–Answer)
RAG (Retrieval-Augmented Generation)
SNOMED-CT (Systematized Nomenclature of Medicine – Clinical Terms)
SO (Specific Objective)
STSE (Short-Term Staff Exchange)
TIA (Transient Ischemic Attack)
TRL (Technology Readiness Level)
TU GRAZ (Graz University of Technology)
UMN (Upper Motor Neuron)
WP (Work Package)
WIDERA (Widening Participation and Spreading Excellence)



Introduction

The methodology presented in this deliverable originates from an initial concept developed in May 2025 during a Short-Term Staff Exchange (STSE), in which four researchers from the Widening coordinator, FCSE, were seconded to the Jožef Stefan Institute (JSI), the project's advanced partner. During this exchange, the FCSE team analyzed a range of neurology educational materials provided by the Widening medical partner FMN, which employ structured, case-based approaches to clinical reasoning.

The objective of this initial work was to derive a generalized, step-by-step reasoning procedure capable of capturing the logic underpinning expert neurological decision-making. This procedure subsequently serves as the conceptual foundation for the design of the Neurology Orchestrator (NeuroOrchestrator), a fine-tuned large language model intended to support and reproduce structured clinical reasoning within real patient care pathways (WP8).

Following the STSE, the methodology was further refined and operationalized within WP7 during the period from September 2025 to January 2026. The resulting pipeline is described in detail in the present deliverable and illustrates a purposefully designed NeuroOrchestrator agentic workflow, spanning case extraction and summarization, hypothesis generation, diagnostic test planning, treatment suggestion, and quality validation.

Building on this methodology, the consortium will construct structured reasoning datasets as part of Deliverable D7.2. These datasets will be used for the fine-tuning and further evaluation of the NeuroOrchestrator, enabling the system to learn from clinically grounded reasoning traces and to emulate expert neurological decision-making in a transparent and verifiable manner.

By applying this methodology, each neurology scenario is ensured to be:

- **Clinically grounded**, derived from real-world neurology workflows as described in established literature;
- **Technically feasible**, clearly representable and suitable for modeling as a downstream large language model task; and
- **Verifiable**, through a purpose-built validation platform in which each scenario includes measurable checkpoints subject to human expert review.

Methodological Approach to Framework Derivation

Neurological decision-making poses specific challenges for AI-supported reasoning, including: (1) a strong reliance on clinical pattern recognition and accurate neuroanatomical localization, (2) frequent diagnostic uncertainty requiring probabilistic and hypothesis-driven thinking, (3) sequential and iterative reasoning as new clinical information emerges, and (4) continuous integration of heterogeneous data from patient history, neurological examination, imaging, and laboratory findings.

Although large language models can generate fluent medical language, they often fail to internalize these structured reasoning processes, motivating the design of datasets that explicitly capture expert neurological logic rather than surface-level textual correlations.

The framework was derived through a systematic review of established neurological diagnostic workflows and an in-depth analysis of expert-authored neurological case descriptions.

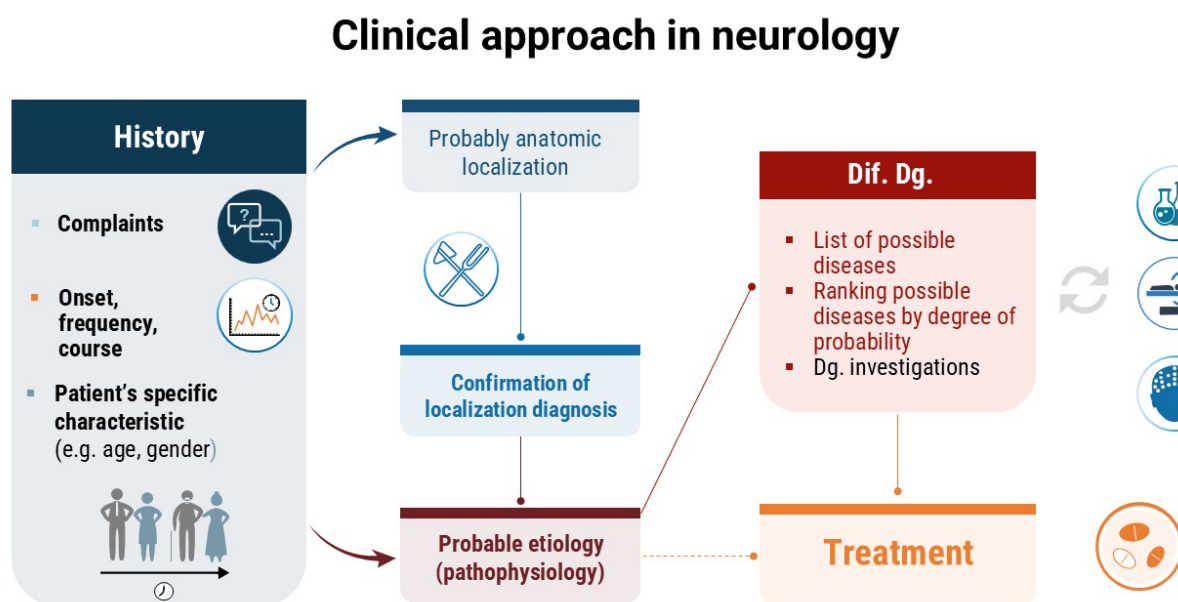
Recurring reasoning steps common across a broad range of neurological disorders were identified and abstracted into a disorder-agnostic reasoning sequence.

This abstraction was iteratively refined through expert input from medically qualified neurologists to ensure clinical relevance and coherence.

The emphasis of this work is on formalizing the structure of clinical reasoning rather than encoding disease-specific guidelines, scoring systems, or treatment protocols.

Clinical Neurology Workflow

This standard neurological clinical approach consists of a series of steps, with primary goals to make anatomic (localization) and pathological (etiology) diagnosis¹ as illustrated on Figure 1. Although standard teaching is that the patient should be allowed to provide history in his or her own words, the process also involves active questioning of the patient to elicit pertinent information and systematic review of previous pertinent medical records.



after Losseff & Scadding (2012)

Figure 1. Clinical neurology approach.

At each step, the neurologist should consider the possible anatomical localizations, the potential pathophysiological mechanisms of disease, and the possible etiologies of the symptoms, especially for the most likely localizations. From the patient's chief complaint and a detailed history, an attentive neurologist can derive clues that lead first to a hypothesis about the location and then to a hypothesis about the etiology of the neurological lesion.²

¹ Losseff N, Scadding J. (editors) Clinical neurology. 4ed. Hodder Arnold 2012

² Jankovic J, Mazziotta JC, Pomeroy SL, Newman NJ (eds.) Bradley and Daroff's Neurology in Clinical Practice. 8th ed. Elsevier; 2021.

From these hypotheses, the experienced neurologist can predict what neurological abnormalities should be present and what should be absent, thereby allowing confirmation of the site of the dysfunction during the neurological examination. Alternatively, analysis of the history may suggest two or more possible anatomical locations and disease mechanisms and etiologies, each with a different predicted constellation of neurological signs. The findings on neurological examination can be used to determine which of these various possibilities is the most likely. After that additional diagnostic tests can be performed to confirm diagnosis. After diagnosis confirmation, adequate treatment plan is designed with respect to specific patients' characteristics (e.g. age, gender, contraindications, allergy) as well as checker for specific issues (e.g. emergency, drug interaction). After treatment implementation monitoring and adjustment of plan is possible as well as iterative hypothesis refinement. For this deliverable, a modified clinical neurology workflow was designed to reflect real world clinical practice while remaining suitable for machine-oriented applications (**Figure 2**).

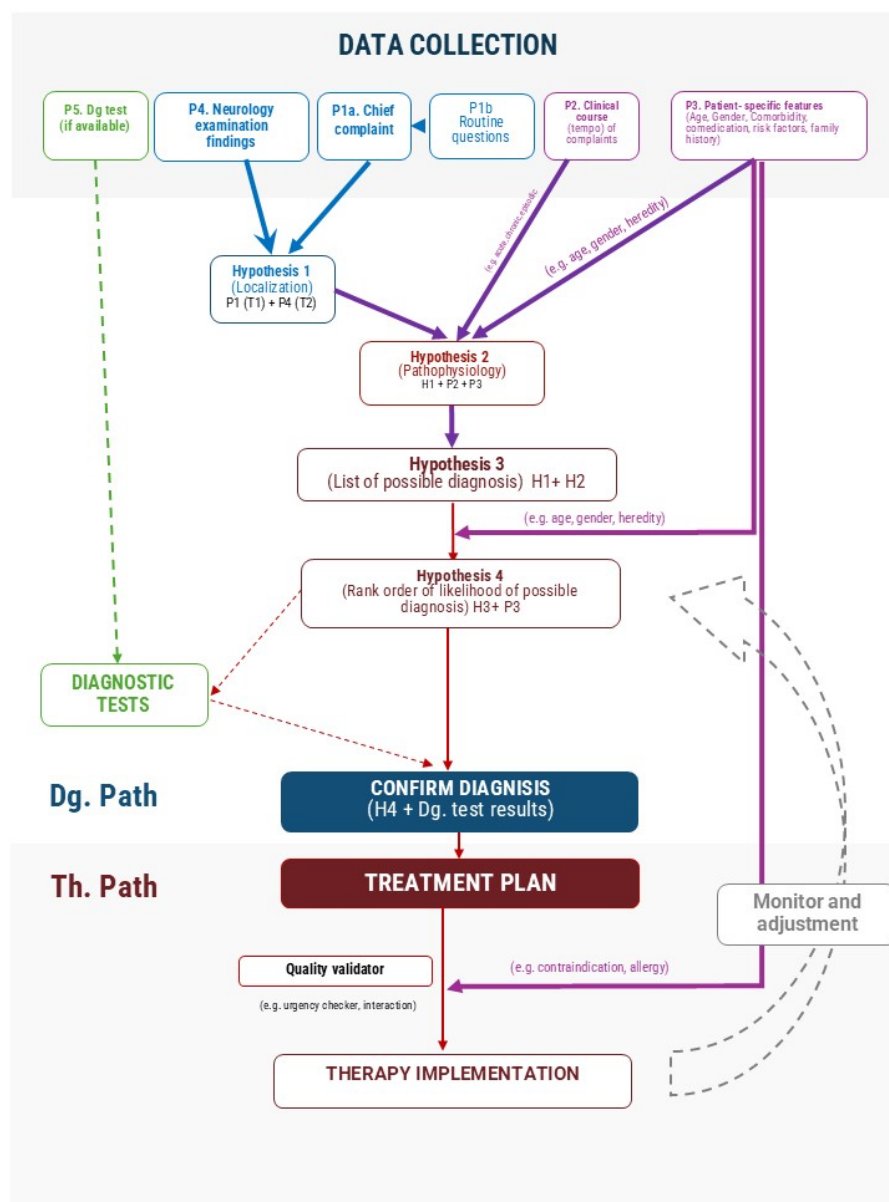


Figure 2. Modified clinical neurology workflow.

The workflow is structured as a sequence of reasoning steps in which the neurologist systematically collects and interprets clinical information with the objective of progressively informing diagnostic and decision-making processes.

The NeuroOrchestrator Agentic Workflow

The NeuroOrchestrator is a modular clinical reasoning pipeline designed to transform raw patient data into a structured and validated neurological decision-support report. The system comprises seven core stages (A0–A7), with an additional safety module (A5.5) embedded within the treatment planning process. Each stage performs a well-defined function and exchanges structured information with other components, enabling hypothesis generation, refinement, validation, and reporting to occur in a coordinated manner. Figure 3 illustrates the complete workflow, and each component is described narratively in the sections that follow.

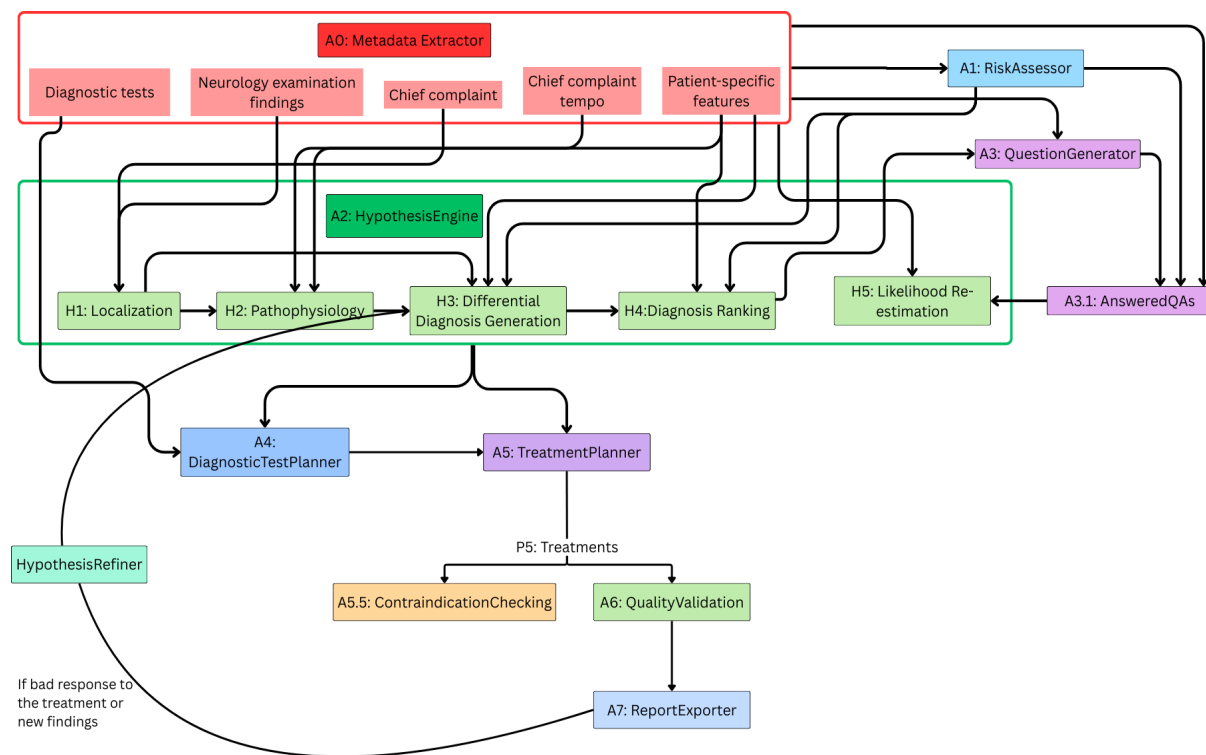


Figure 3. NeuroOchestrator workflow.

Step A0: Case Extraction

The pipeline begins with **A0 (CaseExtractor)**. This agent uses a large language model to extract structured JSON data from unstructured neurological case reports. The extracted information is organized into five main groups: *diagnostic tests* (laboratory, radiology, neurophysiology, and additional investigations), *neurological examination findings*, *chief complaint*, *chief complaint tempo*, and *patient-specific features* (demographics, medical history, medications, allergies, family history, and vital signs). The agent performs strictly



factual extraction without interpretation or clinical reasoning, producing a machine-readable representation that serves as input for downstream agents.

Step A1: Emergency Risk Assessment

A1 (RiskAssessment Agent) performs immediate neurological risk stratification based on the extracted metadata from the case. The agent analyzes clinical findings to detect *neurological red flags* indicative of potentially life-threatening conditions and assigns an overall risk score and urgency level. Identified red flags are explicitly linked to required actions and time windows (e.g., stroke protocol activation, empiric antibiotics, airway protection). In addition, the agent specifies immediate interventions, monitoring requirements, need for specialist consultation, and escalation pathways. The output is a structured JSON object that enables early safety prioritization before further diagnostic reasoning.

Step A2: Hypothesis Generation

The **A2 (Hypothesis Generator)** is the central clinical reasoning component of the NeuroOrchestrator. Its primary task is to transform structured clinical metadata into a coherent, ranked set of diagnostic hypotheses by emulating expert neurological reasoning. Unlike A0 and the RiskAssessment agent, this component performs interpretative reasoning, operating over extracted facts rather than raw text.

The Hypothesis Generator is internally decomposed into a sequence of reasoning sub-steps, each consuming specific inputs and producing structured intermediate outputs:

H1: Localization- detects the most likely **anatomical position(s)** of the lesion based on the chief complaint, chief complaint tempo, and neurological examination findings, producing structured localization hypotheses without assigning specific disease entities.

H2: Pathophysiology- infers the most likely **underlying pathophysiological processes** associated with the previously identified anatomical localizations, using patient-specific features (age, sex, medical and family history, medications) and chief complaint tempo. It produces structured hypotheses describing the type of biological process (e.g., ischemic, inflammatory, degenerative, infectious) without assigning specific disease labels.

H3: Differential Diagnosis Generation- produces a structured list of **specific neurological diagnoses** by combining the inferred anatomical localization(s) and pathophysiological processes with the patient's presentation and risk assessment output. Red flags, urgency level, and overall risk are used to prioritize and filter diagnoses, ensuring that high-severity and time-critical conditions are considered first. The output is a set of plausible diagnoses, each accompanied by a brief clinical rationale, without ranking or assigning probabilities.

H4: Diagnosis Ranking- re-evaluates the unranked differential diagnosis list in the context of patient-specific features and assigns an explicit likelihood score to each diagnosis. The agent orders diagnoses from most to least likely and provides a brief rationale justifying each likelihood.



H5: Likelihood Re-estimation- updates the ranked diagnostic hypotheses by incorporating new evidence obtained from answered clinical questions. Using only existing patient metadata and QA-derived information, this step revises likelihood estimates, reorders diagnoses accordingly, and introduces a limited number of new hypotheses explicitly justified by the QA evidence. The output is an updated, consistently ranked set of diagnostic hypotheses reflecting the refined clinical picture.

Step A3: Question Generation

A3 (Question Generation) identifies information gaps relevant to the top-ranked diagnostic hypotheses and generates structured, hypothesis-driven questions to address those gaps. For each leading hypothesis, it produces targeted core, supporting, and exclusionary questions designed to confirm, refine, or reject competing diagnoses.

Step A3.1: AnsweredQAs

A3.1 (Question Resolution) determines whether each question generated in A3 can be answered using the available clinical text and structured patient metadata. When explicit information is present, the question is resolved by extracting the corresponding answer along with a concise justification and supporting evidence. The resulting set is passed to *H5 (Hypothesis Refinement)*, where it is used to update and re-rank diagnostic hypothesis likelihoods.

Step A4: Test Planner

A4 (TestPlanner) converts hypotheses into a structured diagnostic work-up plan. It recommends imaging, labs, electrophysiology, or specialist referrals. Each test is tied to a rationale and which hypothesis it helps to confirm or exclude.

Step A5: Treatment Planner

A5 takes A0 metadata, A2 hypothesis and A4 tests to produce a preliminary treatment strategy including medications, lifestyle recommendations, and urgency-based triage. Treatments are mapped to the hypothesized conditions and adjusted based on comorbidities extracted in A0.

Step A5.5: Contraindication Checking (Optional)

If A5 proposes any pharmacological options, **A5.5** runs an automated safety screen using extracted medications, age, pregnancy status, allergies, and comorbidities. It outputs warnings or alternative treatment suggestions.

Step A6: Quality Validation

A6 performs structured checks for completeness (required outputs present, likelihood constraints, question–answer coverage), cross-step consistency (alignment between final hypotheses, investigations, treatments, and QA evidence), coverage of critical life-threatening diagnoses, guideline citation/compliance, and safety-net/follow-up clarity. The output is a



machine-readable ValidationReport that lists missing elements, inconsistencies with severity, quality flags, confidence, and whether human review is required (with urgency).

Step A7: Report Export

The pipeline concludes with A7, which synthesizes everything into a clinician-ready structured report (JSON + narrative text) that can be exported via HL7-CDA, FHIR, or internal formats. This is the final product delivered to downstream systems or clinicians.

Hypothesis Refiner

The **Hypothesis Refiner** updates the differential diagnosis as new diagnostic information becomes available, such as laboratory results, imaging findings, or neurophysiological studies. Instead of regenerating diagnoses from scratch, it incrementally revises the existing hypotheses to reflect the accumulating clinical evidence. For each diagnosis, the refiner applies Bayesian reasoning: the current likelihood of a diagnosis is treated as a *pre-test probability*, which is then adjusted based on how strongly new findings support or contradict that diagnosis. Confirmatory findings increase the probability, while contradictory or definitively negative results reduce it. Diagnoses may be ruled out when high-negative-predictive-value tests or absent key criteria make them unlikely, and new diagnoses may be introduced when unexpected findings strongly suggest an alternative explanation. This probabilistic updating is combined with established diagnostic criteria and clinical patterns, ensuring that the refined differential remains clinically coherent and evidence based. The output is an updated set of diagnostic hypotheses with revised likelihoods and a concise summary explaining how the new evidence influenced the diagnostic reasoning. When no new diagnostic information is available, the existing differential is preserved and explicitly reported as unchanged.

Prompt-Based Agentic Mapping of Free-Text Clinical Narratives

The following **Table 1** provides a complete overview of the prompt engineering strategy used across all NeuroOrchestrator steps. Each step is governed by a dedicated SYSTEM prompt that defines its clinical role, reasoning style, constraints, and required output structure. These prompts ensure that the pipeline behaves consistently across steps, maintains clinical correctness, and produces machine-readable outputs aligned with the orchestrator's data model. By standardizing the prompt instructions for every step (from case extraction and summarization to hypothesis generation, risk assessment, question synthesis, diagnostic planning, treatment modeling, and final report export) the NeuroOrchestrator achieves reproducible, interpretable, and domain-specialized behavior. The table below summarizes the full prompt text for each step (A0–A7), serving as the authoritative definition of agent behavior within the system.

Table 1. NeuroOrchestrator prompt engineering strategy.

Task	Prompt
------	--------



A0: CaseExtractor - parse raw upload into structured dict	NOT an LLM prompt (en_core_sci_md (scispacy) may be used: AGE_SEX = r"(\d{1,3})[-]?(year[-]old)?\s+(male female boy girl)" CC_PATTERN = r"chief complaint:? (.*?)[\.\n]" VITALS_TABLE = r"(BP HR RR Temp)[^\n]+" or LLM prompt: SYSTEM: You are an information-extraction engine for neurology case reports. Return ONLY a JSON object with the requested keys. If a key is not found, set its value to null. Do NOT invent data. TASK: Extract the following fields from the clinical text: • age (int, years) • sex ("male" "female") • chief_complaint (string, plain text, include primary presenting symptom as written) • chief_complaint_tempo (string, plain text, include tempo of pattern of the chief complaint) • date_of_event (ISO8601 yyyy-mm-dd or null) • past_medical_history (comma-separated string; max 120 chars, include prior diseases/conditions) • family_history (string; include only neurologically relevant items) • medication_list (string; comma-separated, include all current medications/therapies as written) • vitals_raw (string; copy the whole vitals line/block if present) TEXT """" {{ source_text }} """" OUTPUT FORMAT: <pre>{ "age": int, "sex": str, "chief_complaint": str, "chief_complaint_tempo": str, "date_of_event": str, "past_medical_history": str, "allergies": str, "family_history": str, "medication_list": str, "neuro_exam_findings": str, "psychic_exam": str, "vitals_raw": str, "laboratory_tests": str, "radiology_tests": str, "neurophysiology_tests": str, "additional_tests": str }</pre>
A1: Risk Assessor (Emergency Risk Assessor)	SYSTEM: You are an emergency neurologist performing immediate risk assessment. Your role is to identify life-threatening conditions and determine urgency of intervention. INPUT: Patient Metadata TASK: 1. Assess overall neurological risk (0.0 = stable, 1.0 = critical) 2. Determine urgency level based on:



	- Emergency (immediate, life-threatening): stroke, status epilepticus, meningitis, raised ICP - Urgent (within 24h): first seizure, progressive weakness, severe headache - Routine (scheduled): chronic conditions, stable symptoms 3. Identify all red flags with specific actions and time windows 4. List any immediate interventions needed before further testing 5. Specify monitoring requirements (telemetry, neuro checks frequency) 6. Determine if specialist consultation needed immediately CRITICAL RED FLAGS TO CHECK: - Thunderclap headache -> CT angiography within 1 hour - Decreased GCS -> secure airway, neuroprotective measures - Signs of raised ICP -> immediate neurosurgery consult - Fever + neck stiffness + confusion -> empiric antibiotics before LP - Sudden onset focal deficit -> stroke protocol activation - Status epilepticus features -> benzodiazepines immediately OUTPUT FORMAT: <pre>{ "overall_risk_score": float, "urgency_level": str, "red_flags": [...], "immediate_interventions": str, "monitoring_requirements": str, "specialist_consultation_needed": str, "escalation_pathway": str }</pre>
A2: HypothesisEngine	
H1:Localization	SYSTEM: You are an experienced neurologist. INPUT: Chief complaint: A0.meta.get("chief_complaint") Chief complaint tempo: A0.meta.get("chief_complaint_tempo") Neurological examination findings: A0.meta.get("neuro_exam_findings") TASK: Infer the most likely anatomical localization(s) of the lesion responsible for the patient's presentation. Return a SINGLE JSON object: - `hypotheses` is a list (max 8 items). - Each item is an object with: - anatomical_position (str): Anatomical position of the suspected lesion (e.g., left frontal lobe, right internal capsule, cervical spinal cord, peripheral nerve, neuromuscular junction, etc.). - rationale (str): One-sentence clinical reasoning that links the symptoms and exam findings to this localization. OUTPUT FORMAT: <pre>[{ "anatomical_position":str, "rationale": str }]</pre>
H2: Pathophysiology	SYSTEM: You are an experienced neurologist. INPUT: Patient specific features: A0.meta.get("patient_specific_features") Chief complaint tempo: A0.meta.get("chief_complaint_tempo") H1 hypothesis TASK:



	Your task is to infer the most likely Pathophysiological process corresponding to localizations, for the specific patient. Return a SINGLE JSON object with the fields: - `hypotheses` is a list (max 8 items). Each item is an object with: - pathophysiology (str): Underlying process - rationale (str): One-sentence explanation links the patient-specific data and chief complaint tempo to this pathophysiology. OUTPUT FORMAT: <pre>[{ "pathophysiology ":str, "rationale": str }]</pre>
H3:Differential Diagnosis Generation	SYSTEM: You are an experienced neurologist. INPUT: H1 hypothesis H2 hypothesis RiskAssessment output TASK: Your task is to generate a DIFFERENTIAL DIAGNOSIS LIST: a set of specific neurological diagnoses that could explain this patient's presentation. Red flags MUST influence the differential diagnosis. Always prioritize explaining high-severity and critical red flags first. Urgency_level must bias the list toward diagnoses that match the clinical tempo. Exclude diagnoses that clearly do NOT fit the red flags, risk profile, or tempo. Use risk information only to PRIORITIZE and FILTER — do NOT restate it as a diagnosis. Return a SINGLE JSON object with the field - `hypotheses` is a list (max 8 items). - Each item is an object with: - diagnosis (str): A specific diagnosis (e.g. "migraine without aura", "idiopathic intracranial hypertension", "cervicogenic headache"). - rationale (str): One-sentence explanation that combines: - the symptoms and tempo, - localization hypotheses, - pathophysiological processes, - and relevant risk factors (age, sex, past/family history, medications). OUTPUT FORMAT: <pre>[{ "diagnosis ":str, "rationale": str }]</pre>
H4: Diagnosis Ranking	SYSTEM: You are an experienced neurologist. INPUT: A0 output AnsweredQAs output TASK: - Re-assess diagnoses for the specific patient. - Rank them from MOST to LEAST likely. - Assign a numeric likelihood (0–1) to each item. Return a SINGLE JSON object with fields:. - `hypotheses` is a list (max 8 items), ordered from MOST to LEAST likely. - Each item is an object with:



	- diagnosis (str) - rationale (str): One-sentence reasoning that explicitly uses patient features (age, sex, tempo, PMH, family_history, medications) to justify the likelihood. - likelihood (float): Probability between 0 and 1. OUTPUT FORMAT: <pre>[{ "diagnosis ":str, "rationale": str, "likelihood": float }]</pre>
H5: Likelihood Re-estimation	SYSTEM: You are an experienced neurologist. INPUT: Patient specific features: A0.meta.get("patient_specific_features") H3 hypothesis TASK: Your task is to update (not reinvent) the differential diagnosis list based on the available question–answer evidence. -Re-assess existing diagnostic hypotheses using the provided QA evidence. -Update and re-rank diagnoses by revising their likelihoods where appropriate. -You may introduce up to two new diagnoses only if strongly supported by QA evidence, and you must explicitly reference the motivating evidence. -Exclude diagnoses that are no longer supported by the evidence or contradict key patient features. -Use only the provided inputs; do not introduce new facts or assumptions. OUTPUT FORMAT: <pre>[{ "diagnosis ":str, "rationale": str, "likelihood": float }]</pre>
A3: QuestionGenerator	SYSTEM: You are an experienced consultant neurologist preparing a structured, history checklist. Output must be valid JSON – no extra keys. INPUT: A0 output RiskAssessment output H4 hypothesis TASK: -Select the TOP 3 hypotheses by likelihood (highest first). -For EACH of the 3 hypotheses, generate EXACTLY 6 questions: - 2 CORE questions - 2 SUPPORTING questions - 2 AGAINST questions TOPIC CONSTRAINT: Assign each question a topic chosen from this ordered list: 01. Cognitive and Emotional Function 02. Seizure and Consciousness History 03. Pain and Headaches 04. Sensory Function 05. Sensory Disturbances 06. Motor Function 07. Fine Motor Skills / Coordination



	08. Gait and Balance 09. Autonomic Function - Order the full list of questions to follow the natural flow of a real neurological history/exam using the topic order above. - Ensure you still meet the exact per-hypothesis counts: 2 core + 2 supporting + 2 against. Return valid JSON including this fields: - question (str) - topic (str) - target (str) - terminology (str or null): SNOMED-CT or LOINC code only, or null if no exact match - hypothesis (str): must match one of the selected top-3 diagnosis names exactly - previous_likelihood: likelihood of hypothesis OUTPUT FORMAT: <pre>[{ "question": str, "topic": str, "target": str, "terminology":str, "hypothesis": str, "previous_likelihood": float }]</pre> STRICT RULES: * Do not invent terminology codes; put ``code": null` if uncertain. * No duplicate questions.
A3.1: AnsweredQAs	SYSTEM: You are a neurologist assistant doing STRICT information extraction. INPUT: A0 output RiskAssessment output A3 questions TASK: For each question, answer using ONLY the information present in the inputs. For each item, return: - diagnosis (copy exactly) - previous_likelihood (copy exactly) - question (copy exactly) - answer: answer to the question - short_justification: one short sentence grounded in the inputs (or null if unknown) - evidence: a short exact phrase from the inputs that supports the answer (or null if unknown) OUTPUT FORMAT: <pre>{ "diagnosis": str, "previous_likelihood": float, "question ": str, "answer":str], "short_justification": str, "evidence": str }</pre>
A4: TestPlanner - produce P4	You are drafting an evidence-based investigation plan. INPUT: A2 hypothesis Diagnostic_tests: A0.meta.get("diagnostic_tests") TASK:



	Propose test/examination/investigation that can confirm or rule out top hypotheses. Each proposed test/examination/investigation should be aligned with the reason why it is needed. Group tests that must be ordered simultaneously, marked with the same "batch_id". Mark each as mandatory/optional and urgent/routine with timing. Make sure that the urgency and necessity are in accordance with international guidelines (e.g., EAN, AAN, NICE). Attach guideline references (e.g., "NICE CG137 1.6.4"). If there are multiple conflicting or consistent recommendations, provide a reference to all of them. You will also receive the following fields: - "laboratory_tests": (string or null) — tests that have ALREADY been completed from the laboratory domain. - "radiology_tests": (string or null) — imaging tests that have ALREADY been performed (MRI, CT, X-ray, ultrasound). - "neurophysiology_tests": (string or null) — neurophysiology investigations ALREADY completed (EEG, EMG, NCS, evoked potentials). - "additional_tests": (string or null) — any other diagnostic tests already performed (genetic tests, CSF analysis, biopsies, etc.) These fields list investigations that are already completed so that you DO NOT recommend them again. Only propose NEW investigations that have not yet been carried out and are needed to confirm or rule out the working hypotheses. For every investigation include: - "investigation": (str) - The test/examination/investigation that the patient should make. - "batch_id": (int) - The order id of the investigation. Lower id means more urgent investigation. Set the same batch id if the investigations should be performed simultaneously. - "guideline_refs": (str)— Attach guideline references (e.g., "NICE CG137 1.6.4"). If there are multiple conflicting or consistent recommendations, provide a reference to all of them. - "reason": (str) — The reason why the patient should take the investigation/test. - "urgency": (str) - Urgent or non-urgent - "mandatory": (str) - Is the investigation mandatory or optional? - "timing": (str) - The timeframe within which the patient should perform the investigation/test. OUTPUT FORMAT: <pre>[{ "investigation": str, "batch_id": str, "guideline_refs":str, "reason": str, "urgency": str, "mandatory": str, "timing": str }]</pre>
A5: TreatmentPlanner - produce P5	You are a neurologist that prepares an initial therapy plan. INPUT: A0 metadata A2 hypothesis A4 treatments TASK: For the highest ranked hypothesis, list suitable therapy or "watchful waiting".



	• Link to hypothesis by index. • Include dosage, timing, and brief reason citing guideline sections. OUTPUT FORMAT: <pre>[{ "linked_hypothesis": int, "name": str, "dose": str, "timing": str, "reason": str }]</pre>
A5.5: ContraindicationChecker - Contraindication and Drug Interaction Checker (Clinical Pharmacist)	SYSTEM: You are a clinical pharmacist specializing in neurology reviewing medication safety and interactions. INPUT: A5 Proposed Treatments A0 Metadata TASK: 1. Check ALL drug-drug interactions: - Between proposed and current medications - Between multiple proposed medications - Classify severity: contraindicated > major > moderate > minor 2. Identify contraindications: - Absolute (never give) - Relative (use with caution) - Provide alternatives for contraindicated drugs 3. Calculate dose adjustments: - Elderly (>65): consider reduced dosing - Renal impairment: use CrCl if available - Hepatic impairment: check LFTs - Pediatric: weight-based dosing 4. Check for therapeutic duplications 5. Create monitoring plan for high-risk drugs Critical interactions to check: - Enzyme inducers (carbamazepine, phenytoin) → reduced efficacy of other drugs - Enzyme inhibitors → increased toxicity - QT prolongation risk (multiple psychotropics) - Bleeding risk (antiplatelet + anticoagulant) - Serotonin syndrome risk (SSRIs + other serotonergic drugs) - Lithium interactions (NSAIDs, ACE-I, thiazides) Neurological drug considerations: - Antiepileptics: check for Stevens-Johnson risk (HLA-B*1502 in Asians) - Benzodiazepines in elderly: fall risk, cognitive impairment



	- Anticholinergics: cognitive burden in elderly - Dopamine antagonists: extrapyramidal symptoms OUTPUT FORMAT: <pre>{ "drug_interactions": [...], "contraindications": [...], "dose_adjustments": [...], "therapeutic_duplications": [...], "monitoring_plan": [...], "safe_to_proceed": bool, "requires_modification": bool, "pharmacy_consultation_needed": bool, "high_alert_medications": [...] }</pre>
A6: QualityValidator - Quality and Completeness Validator (Clinical Audit System)	SYSTEM: You are a clinical quality auditor validating the completeness and safety of a neurology diagnostic report. INPUT: - A0: patient metadata (history/exam and available tests lists) - A1: risk assessment output (urgency, red flags, escalation pathway, etc.) - H1: localization hypotheses - H2: pathophysiology hypotheses - H3: unranked diagnosis list (generated differential) - H4: ranked diagnosis hypotheses BEFORE question-answer refinement - H5: ranked diagnosis hypotheses AFTER incorporating answers (updated differential) - A3: question plan for top diagnoses (questions have question_type: core/supporting/against) - A3.1: answered questions (yes/no/unknown + optional evidence) - A4: investigation plan (labs/imaging/neurophysiology/etc.) - A5: treatments plan - A5.5: safety net / patient instructions / follow-up guidance VALIDATION CHECKLIST: 1) Completeness A) Required outputs_csv present and non-empty: - h5 (final ranked hypotheses) - a3 (questions) - a3.1 (answers) - a4 (investigations) - a5 (treatments) - a5.5 (safety net / follow-up) - a1 (risk assessment) B) Hypotheses quality checks: - a2.hypotheses length <= 8 - each likelihood within [0,1] - strictly sorted descending by likelihood - sum of likelihoods <= 1.0 (allow tolerance up to 1.01) C) Question plan checks (based on what a3 actually contains): - Determine TOP 3 hypotheses from h4 (highest likelihood). - Confirm that a3 contains at least 6 questions PER each of those top-3 hypotheses. - Confirm every a3 question has: question, topic, target; terminology may be null. - Confirm a3.1 provides an answered entry for every a3 question by matching: - question text (exact or near-exact) - and hypothesis string (a3.hypothesis should match a3.1.diagnosis when possible)



	If matches are missing, list them in missing_elements. D) Investigations/treatments linkage: - a4 investigations should be plausibly relevant to the TOP hypotheses in a2. - a5 treatments should be plausibly related to hypotheses and not obviously unsafe given a0 + a1. 2) Consistency (limited by available fields) - Since questions are NOT labeled core/supporting/against, DO NOT apply core/against-specific logic. - Instead check: - If a3.1 contains evidence/justifications that strongly contradict the top diagnosis in H5 (e.g., clear denial of defining features), but H5 keeps it unchanged from H4, flag "Update Not Responsive". - If H5 is identical to H4 despite a3.1 containing multiple non-empty answers/evidence, flag "Refinement Ineffective" (minor/major depending on case). - Ensure a4 and a5 align with final H5 more than stale H4. 3) Critical diagnoses coverage (life-threatening) Verify whether the workflow considered or ruled out, when clinically plausible from a0/a1: - Stroke/TIA - Meningitis/encephalitis - Ongoing seizure/status epilepticus - Raised intracranial pressure - Spinal cord compression - Guillain-Barré syndrome - Myasthenia gravis crisis If red flags/high urgency exist in a1 and these are not reflected in hypotheses or investigations, add to missed_critical. 4) Guideline compliance (strict) - Count ONLY explicit guideline identifiers that appear in the provided artifacts (e.g., "NICE CG186", "AAN", "EAN"). - If no explicit guideline strings are present, set guidelines_cited = 0. - If the plan implies guideline-governed decisions but has zero explicit citations, list likely missing citations. 5) Safety net + follow-up - Confirm a5.5 contains emergency/red-flag advice and a follow-up plan. - Cross-check a1 escalation pathway is not contradicted by a5.5. OUTPUT FORMAT: <pre>{ "is_complete": bool, "completeness_score": float, "missing_elements": list[str], "inconsistencies": list[str], "critical_diagnoses_check": dict, "guideline_compliance": dict, "safety_net_included": bool, "patient_instructions_clear": bool, "follow_up_specified": bool, "confidence_score": float, "requires_human_review": bool, "review_urgency": str, "quality_flags": list }</pre>
A7: ReportExporter - produce final output in	Raw text -> A0 -> meta -> A1 -> RiskAssessment



a specified format (json, xlsx...)	-> A2 ->Hypothesis -> A3 -> Questions ->A3.1->Answers -> A4 -> Investigations -> A5 -> Treatments -> A6 -> .xlsx / .json
Hypothesis Refiner	You are a differential diagnosis expert capable of refining differential diagnoses. You may consider using Bayesian probability and clinical decision theory. You update diagnostic probabilities based on new evidence (diagnostic results) while considering test characteristics and disease prevalence. INPUT: Current Differential: {hypotheses} All Previous Results: {diagnostic_results} Clinical Trajectory: {clinical_trajectory} TASK: - Bayesian update for each diagnosis (make deeper research about this): - Pre-test probability: current probability - Likelihood ratio: based on test sensitivity/specificity - Post-test probability: calculated update - Consider disease prevalence (base rates) - Diagnostic criteria check: - For each diagnosis, check if formal criteria are met - Examples: * McDonald criteria for MS * ICHD-3 for headache disorders - Rule out diagnoses if: - Definitive negative test (NPV >95%) - Multiple contradicting findings - Probability <0.05 - Key criterion definitively absent - Rule in diagnoses if: - Pathognomonic finding present - All major criteria met - Probability >0.90 with confirming test - Add new diagnoses if: - Unexpected findings suggest unconsidered diagnosis - Clinical trajectory doesn't match current differential - Incidental finding requires investigation - Assess convergence: - Strong convergence: top diagnosis >0.80, others <0.20 - Moderate: top diagnosis 0.50-0.80 - Weak: all diagnoses <0.50 - Divergent: findings support multiple diagnoses equally - Identify next test: - What would maximally discriminate between top 2 diagnoses? - Consider: sensitivity, specificity, cost, risk, availability Special considerations: - avoid sticking to initial diagnosis OUTPUT FORMAT: <pre>{ [{ "anatomical_position": "string", "pathophysiology": "string", "rationale": "string", "likelihood": 0.0 }], }</pre>



	"differential_summary": "string"
	}

The following Table 2 specifies the complete JSON output schemas expected from each NeuroOrchestrator step/agent. These schemas define the structure, required fields, data types, and semantic roles of every output object produced throughout the pipeline. By enforcing a precise and machine-readable format the NeuroOrchestrator ensures interoperability across agents, supports downstream validation and quality checks, and enables seamless integration with external systems such as HL7-CDA, FHIR, or hospital information platforms.

Table 2. Stepwise JSON schemas.

Task	JSON Schema
A0: CaseExtractor - parse raw upload into structured dict	<pre> meta: dict = { "age": int None, "sex": str None, # "male" or "female" "chief_complaint": str None, "chief_complaint_tempo": str None, "date_of_event": str None, # ISO8601 yyyy-mm-dd "past_medical_history": str None, "allergies": str None, "family_history": str None, "medication_list": str None, "neuro_exam_findings": str None, "psychic_exam": str None, "vitals_raw": str None, "laboratory_tests": str None, "radiology_tests": str None, "neurophysiology_tests": str None, "additional_tests": str None } </pre>
A1: Risk Assessor (Emergency Risk Assessor)	<pre> risk_assessment: dict = { "overall_risk_score": float, "urgency_level": string("emergency" "urgent" "routine"), "red_flags": [{ "finding": string, "severity": string("critical" "high" "moderate" "low"), "action_required": string, "time_window": string }], "immediate_interventions": string, "monitoring_requirements": string, "specialist_consultation_needed": string, "escalation_pathway": string } </pre>
A2: HypothesisEngine	
H1: Localization	<pre> localization_hypotheses: dict = { "hypotheses": [{ "anatomical_position": string, "rationale": string }] } </pre>



H2: Pathophysiology	<pre> pathophysiology_hypotheses: dict = { "hypotheses": [{ "pathophysiology": string, "rationale": string }] } </pre>
H3: Differential Diagnosis Generation	<pre> diagnosis_hypotheses: dict = { "hypotheses": [{ "diagnosis": string, "rationale": string }] } </pre>
H4: Diagnosis Ranking	<pre> ranked_diagnosis_hypotheses: dict = { "hypotheses": [{ "diagnosis": string, "rationale": string, "likelihood": float }] } </pre>
H5: Likelihood Re-estimation	<pre> reranked_diagnosis_hypotheses: dict = { "hypotheses": [{ "diagnosis": string, "rationale": string, "likelihood": float }] } </pre>
A3: QuestionGenerator	<pre> questions_output: dict = { "questions": [{ "question": string, "topic": string, "target": string, "terminology": string null, "hypothesis": string, "previous_likelihood": float null }] } </pre>
A3.1: AnsweredQAs	<pre> answered_qas_output: dict = { "qa": [{ "diagnosis": string null, "previous_likelihood": float null, "question": string null, "answer": string null, "short_justification": string null, "evidence": string null }] } </pre>



	} }
A4: TestPlanner - produce P4	<pre> investigation_plan_output: dict = { "investigations": [{ "investigation": string, "batch_id": int, "guideline_refs": string, "reason": string, "urgency": string, # e.g., "urgent" "non-urgent" "mandatory": string, # e.g., "mandatory" "optional" "timing": string }] } </pre>
A5: TreatmentPlanner - produce P5	<pre> treatment_plan_output: dict = { "treatments": [{ "treatment": string null, "dose": string null, "timing": string null, "reason": string null, "linked_hypothesis": string null, "mandatory": string null }] } </pre>
A5.5: ContraindicationChecker - Contraindication and Drug Interaction Checker	<pre> safety_check_output: dict = { "safety_check": { "drug_interactions": [{ "drug1": string, "drug2": string, "severity": string, # "contraindicated" "major" "moderate" "minor" "mechanism": string, "clinical_effect": string, "management": string # "avoid" "adjust_dose" "monitor" "separate_timing" }], "contraindications": [{ "drug": string, "reason": string, "type": string, # "absolute" "relative" "alternative_options": [string] }], "dose_adjustments": [{ "drug": string, "factor": string, # "age" "weight" "renal" "hepatic" "original_dose": string, "adjusted_dose": string, "rationale": string, "monitoring_required": [string] }] }, } </pre>



	<pre> "therapeutic_duplications": [string], "monitoring_plan": [{ "parameter": string, "frequency": string, "target_range": string }], "allergy_alerts": [{ "substance": string, "reaction": string, "severity": string, # "mild" "moderate" "severe" "anaphylaxis" "management": string }], "safe_to_proceed": bool, "requires_modification": bool, "pharmacy_consultation_needed": bool, "high_alert_medications": [string] } } </pre>
A6: QualityValidator - Quality and Completeness Validator (Clinical Audit System)	<pre> validation_report: dict = { "validation_report": { "is_complete": bool, "completeness_score": float, "missing_elements": [string], "inconsistencies": [{ "type": string, "description": string, "severity": string ("critical" "major" "minor") }], "critical_diagnoses_check": { "all_life_threatening_considered": bool, "missed_critical": [string] }, "guideline_compliance": { "guidelines_cited": int, "compliance_rate": float, "missing_citations": [string] }, "safety_net_included": bool, "patient_instructions_clear": bool, "follow_up_specified": bool, "confidence_score": float, "requires_human_review": bool, "review_urgency": string ("immediate" "urgent" "routine"), "quality_flags": [string] } } </pre>
A7: ReportExporter - produce final output in a specified format (json, xlsx...)	<pre> export_descriptor: dict = {...} </pre>
Hypothesis Refiner	<pre> refined_hypotheses: dict = { [{ </pre>



	<pre> "anatomical_position": "string", "pathophysiology": "string", "rationale": "string", "likelihood": 0.0 }], "differential_summary": "string" } </pre>
--	---

The Neurology Scenarios Mapping

The NeuroOrchestrator methodology is designed to operate across the full spectrum of neurological presentation types, from acute, time-critical emergencies to complex chronic conditions requiring longitudinal reasoning. To validate the generality and robustness of the pipeline, representative raw-text clinical narratives were drawn from established neurology case-study collections that mirror real-world diagnostic diversity. These include works in epilepsy, stroke medicine, general neurology, and classic foundational descriptions of neurological syndromes (e.g., PubMed-indexed case reports; *Case Files Neurology*, 3rd ed. (2018); *Neurology: Clinical Cases Uncovered*, (2011); *Clinical Cases in Neurology*, 2nd ed. (2024); *Fifty Neurologic Cases from Mayo Clinic*, (2004)). These sources provide heterogeneous, narrative-rich patient descriptions containing symptoms, timelines, risk factors, examination findings, and investigative data, therefore allowing the Orchestrator's structured extraction prompts to be evaluated in realistic conditions. To illustrate how the methodology processes raw clinical narratives, one example was processed through the pipeline.

Neurology Scenario Example

Case: A 65-year-old man was evaluated for a 5-year history of a slowly progressive gait difficulty that was subacute in onset. He complained of stiffness of the lower limbs and difficulty with balance. Three years after the onset of gait difficulty, he began to use a cane and 1 year later, to use a walker. For 1 year before evaluation, he had paresthesias of the hands and feet. He also had a 5-year history of erectile dysfunction and a 1-year history of increased urinary frequency with urgency and hesitancy. His medications included aspirin and ranitidine. For the past 22 years, he had been taking 200 to 400 mg of zinc daily to prevent colds. Neurologic examination demonstrated increased tone and mild weakness in the lower limbs, with an upper motor neuron pattern. He had difficulty on the heel-to-shin test. Rapid foot tapping was slow and clumsy. Vibration sensation in the lower limbs was impaired in a graded manner up to the knees, and position sense was impaired at the toes and malleoli. The reflexes were brisk except the ankle jerk, which was slightly depressed. The plantar response was extensor. He had a spastic-ataxic gait and was unable to walk tandemly. The Romberg test was positive.³

³ Noseworthy, J. H. (Ed.). (2004). *Fifty neurologic cases from Mayo Clinic* (1st ed.). Oxford University Press.

**Table 3.** Neurology scenario example.**A0:**

```
{
  "age": 65,
  "sex": "male",
  "chief_complaint": "5-year history of a slowly progressive gait difficulty",
  "neuro_exam_findings": "Neurologic examination demonstrated increased tone and mild weakness in the lower limbs, with an upper motor neuron pattern. He had difficulty on the heel-to-shin test. Rapid foot tapping was slow and clumsy. Vibration sensation in the lower limbs was impaired in a graded manner up to the knees, and position sense was impaired at the toes and malleoli. The reflexes were brisk except the ankle jerk, which was slightly depressed. The plantar response was extensor. He had a spastic-ataxic gait and was unable to walk tandemly. The Romberg test was positive.",
  "chief_complaint_tempo": "subacute onset, slowly progressive over five years",
  "date_of_event": null,
  "past_medical_history": "erectile dysfunction (5 years), urinary frequency/urgency/hesitancy (1 year), paresthesias in hands and feet (1 year before evaluation)",
  "allergies": null,
  "family_history": null,
  "medication_list": "aspirin, ranitidine, zinc (200 to 400 mg daily for 22 years)",
  "psychic_exam": null,
  "vitals_raw": null,
  "laboratory_tests": null,
  "radiology_tests": null,
  "neurophysiology_tests": null,
  "additional_tests": null
}
```

A1:

```
{
  "overall_risk_score": 0.2,
  "urgency_level": "routine",
  "red_flags": [
    {
      "finding": "Chronic, progressive spastic paraparesis with upper motor neuron signs, impaired posterior column function, and a positive Romberg test",
      "severity": "moderate",
      "action_required": "Further workup for subacute combined degeneration, cervical myelopathy, or other progressive myelopathies (including MRI spine and serum B12/copper/zinc levels). Consider endocrine and metabolic etiologies. Assess for potential zinc toxicity due to chronic supplement use.",
      "time_window": "Refer to neurology within 1-2 weeks; expedite if symptoms worsen."
    },
    {
      "finding": "Bladder dysfunction and erectile dysfunction in the context of progressive myelopathy",
      "severity": "moderate",
      "action_required": "Assess for urinary retention and risk of infection; consider urology referral. Monitor for any acute worsening.",
      "time_window": "Outpatient assessment within 1-2 weeks; escalate if new retention or incontinence develops."
    }
  ],
  "immediate_interventions": "No immediate interventions required at this time; patient is currently neurologically stable.",
  "monitoring_requirements": "Routine outpatient monitoring. No continuous telemetry or frequent neuro checks required.",
  "specialist_consultation_needed": "Outpatient neurology evaluation for further workup (e.g., neuroimaging, laboratory studies for myelopathy or metabolic causes).",
  "escalation_pathway": "If patient develops acute neurological deterioration (sudden onset weakness, new sensory level, incontinence, confusion), refer to emergency department immediately for urgent assessment."
}
```

A2-H1:

```
{
  "hypotheses": [
    {
      "anatomical_position": "thoracic spinal cord (dorsal columns and corticospinal tracts)",
      "rationale": "Spastic paraparesis, impaired vibration and position sense, brisk reflexes with extensor plantar"
    }
  ]
}
```



responses indicate combined corticospinal and posterior column involvement affecting the thoracic cord."

```

    },
    {
      "anatomical_position": "cervical spinal cord (dorsal columns and corticospinal tracts)",
      "rationale": "Bilateral lower limb UMN signs and posterior column dysfunction may arise from a cervical cord process, though arm involvement is minimal."
    },
    {
      "anatomical_position": "midline dorsal spinal cord",
      "rationale": "A midline lesion involving dorsal columns and corticospinal tracts explains sensory ataxia with spasticity."
    }
  ]
}

```

A2-H2:

```

{
  "hypotheses": [
    {
      "pathophysiology": "chronic metabolic or nutritional neurotoxicity",
      "rationale": "Long-standing excessive zinc intake can induce copper deficiency, leading to progressive degeneration of spinal cord long tracts."
    },
    {
      "pathophysiology": "slowly progressive degenerative myelopathy",
      "rationale": "The indolent, bilateral, length-dependent progression over years supports a chronic degenerative process."
    },
    {
      "pathophysiology": "chronic compressive myelopathy",
      "rationale": "Slow progression with bladder involvement raises consideration of long-standing spinal cord compression."
    }
  ]
}

```

A2-H3:

```

{
  "hypotheses": [
    {
      "diagnosis": "Subacute combined degeneration (B12 or copper deficiency)",
      "rationale": "The chronic progression of spastic paraparesis, impaired posterior column function, positive Romberg sign, and urinary and erectile dysfunction in a 65-year-old man with a history of zinc overuse supports a myelopathy due to B12 or copper deficiency, as excess zinc can induce copper deficiency and neurologically mimic B12 deficiency."
    },
    {
      "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
      "rationale": "Long-term high-dose zinc supplementation can lead to copper deficiency, resulting in progressive myelopathy and peripheral neuropathy with combined dorsal column and corticospinal tract involvement, closely matching the patient's symptoms and disease tempo."
    },
    {
      "diagnosis": "Cervical spondylotic myelopathy",
      "rationale": "Chronic progressive spastic paraparesis, loss of proprioception, brisk reflexes, and urinary symptoms are typical of cervical spinal cord compression due to degenerative spondylotic changes in older adults."
    },
    {
      "diagnosis": "Primary progressive multiple sclerosis",
      "rationale": "A slowly progressive myelopathy with upper motor neuron and posterior column features, along with sphincter dysfunction, may reflect a primary progressive demyelinating disorder affecting the spinal cord in late adulthood."
    },
    {
      "diagnosis": "HTLV-1-associated myelopathy (tropical spastic paraparesis)",
      "rationale": "A chronic, slowly progressive spastic paraparesis with urinary symptoms evolving over years can

```



be seen in HTLV-1-associated myelopathy, particularly in individuals with relevant epidemiological exposure."

```

    },
    {
      "diagnosis": "Hereditary spastic paraplegia (adult-onset form)",
      "rationale": "Progressive lower limb spasticity and weakness with bladder involvement over years may be consistent with adult-onset hereditary spastic paraplegia, although late presentation and sensory involvement are less typical."
    },
    {
      "diagnosis": "Spinal dural arteriovenous fistula",
      "rationale": "A chronic, progressive myelopathy with paraparesis, sensory changes, and bladder dysfunction may rarely be due to a spinal dural arteriovenous fistula and should be excluded given the indolent course."
    },
    {
      "diagnosis": "Tabes dorsalis (neurosyphilis)",
      "rationale": "Progressive dorsal column dysfunction, gait instability, and urinary or erectile symptoms in an older adult may be due to late neurosyphilis, particularly in the absence of acute inflammatory features."
    }
  ]
}

```

A2-H4:

```

{
  "hypotheses": [
    {
      "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
      "rationale": "The 65-year-old man's long-term high-dose zinc supplementation (200–400 mg daily for 22 years), progressive spastic paraparesis, posterior column dysfunction, distal paresthesias, and bladder and erectile dysfunction strongly suggest copper deficiency myeloneuropathy secondary to zinc toxicity.",
      "likelihood": 0.5
    },
    {
      "diagnosis": "Subacute combined degeneration (B12 or copper deficiency)",
      "rationale": "Chronic progressive myelopathy with impaired posterior column function and risk factors for copper deficiency due to zinc overuse closely mimics the clinical syndrome of subacute combined degeneration from either B12 or copper deficiency in this older man.",
      "likelihood": 0.2
    },
    {
      "diagnosis": "Cervical spondylotic myelopathy",
      "rationale": "Progressive spastic paraparesis, upper motor neuron signs, and urinary symptoms in a 65-year-old man are compatible with cervical cord compression, though the lack of neck pain and strong dietary exposure history make this less likely than a metabolic etiology.",
      "likelihood": 0.11
    },
    {
      "diagnosis": "Primary progressive multiple sclerosis",
      "rationale": "A chronic myelopathy with both motor and sensory involvement could fit primary progressive multiple sclerosis, but the patient's age and characteristic zinc exposure history make this diagnosis less likely.",
      "likelihood": 0.07
    },
    {
      "diagnosis": "Spinal dural arteriovenous fistula",
      "rationale": "Very slowly progressive myelopathy may be caused by a spinal dural arteriovenous fistula in older men with long-standing gait and bladder disturbances, but the absence of acute deterioration and the exposure history lower its probability.",
      "likelihood": 0.04
    },
    {
      "diagnosis": "HTLV-1-associated myelopathy (tropical spastic paraparesis)",
      "rationale": "The clinical tempo and progressive spastic paraparesis with bladder involvement could be compatible with HTLV-1-associated myelopathy, but there is no evidence of relevant epidemiological exposure.",
      "likelihood": 0.03
    },
    {
      "diagnosis": "Hereditary spastic paraplegia (adult-onset form)",

```



```

"rationale": "Although adult-onset hereditary spastic paraplegia can cause slowly progressive spastic
paraparesis, the absence of a family history and the presence of sensory involvement and zinc exposure make this
diagnosis unlikely.",
"likelihood": 0.03
},
{
"diagnosis": "Tabes dorsalis (neurosyphilis)",
"rationale": "Progressive proprioceptive loss and bladder dysfunction could suggest neurosyphilis, but the lack
of risk factors, more prominent corticospinal signs, and clear iatrogenic exposure history make this the least likely
diagnosis.",
"likelihood": 0.02
}
]
}

```

A3:

```

{
"questions": [
{
"question": "Have you noticed any changes in your mood, memory, or ability to think clearly over the past
several years?",
"topic": "Cognitive and Emotional Function",
"target": "patient",
"terminology": null,
"linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
"previous_likelihood": 0.5
},
{
"question": "Have you experienced any episodes of confusion or unusual emotions recently?",
"topic": "Cognitive and Emotional Function",
"target": "patient",
"terminology": null,
"linked_hypothesis": "Subacute combined degeneration (B12 or copper deficiency)",
"previous_likelihood": 0.2
},
{
"question": "Has anyone in your family had similar problems with movement or cognition?",
"topic": "Cognitive and Emotional Function",
"target": "patient",
"terminology": null,
"linked_hypothesis": "Cervical spondylotic myelopathy",
"previous_likelihood": 0.11
},
{
"question": "Have you ever had episodes of passing out, loss of consciousness, or seizure-like activity?",
"topic": "Seizure and Consciousness History",
"target": "patient",
"terminology": "10149003",
"linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
"previous_likelihood": 0.5
},
{
"question": "Have you ever lost consciousness suddenly or felt faint?",
"topic": "Seizure and Consciousness History",
"target": "patient",
"terminology": "271327008",
"linked_hypothesis": "Subacute combined degeneration (B12 or copper deficiency)",
"previous_likelihood": 0.2
},
{
"question": "Have you had any episodes of sudden falls without warning?",
"topic": "Seizure and Consciousness History",
"target": "patient",
"terminology": null,
"linked_hypothesis": "Cervical spondylotic myelopathy",
"previous_likelihood": 0.11
}
]
}

```



```

    },
    {
      "question": "Do you experience any numbness, tingling, burning, or unusual sensations in your legs or hands?",
      "topic": "Sensory Disturbances",
      "target": "patient",
      "terminology": "162469003",
      "linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
      "previous_likelihood": 0.5
    },
    {
      "question": "Have you noticed any recent difficulties feeling vibrations or knowing where your feet are?",
      "topic": "Sensory Function",
      "target": "patient",
      "terminology": "271442005",
      "linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
      "previous_likelihood": 0.5
    },
    {
      "question": "Have you experienced pins-and-needles sensations or loss of feeling in your hands or feet?",
      "topic": "Sensory Disturbances",
      "target": "patient",
      "terminology": "24700007",
      "linked_hypothesis": "Subacute combined degeneration (B12 or copper deficiency)",
      "previous_likelihood": 0.2
    },
    {
      "question": "Have you observed any gradual loss of vibration or position sense in your toes or fingers?",
      "topic": "Sensory Function",
      "target": "patient",
      "terminology": "271442005",
      "linked_hypothesis": "Subacute combined degeneration (B12 or copper deficiency)",
      "previous_likelihood": 0.2
    },
    {
      "question": "Have you felt muscle weakness, heaviness, or difficulty moving your legs?",
      "topic": "Motor Function",
      "target": "patient",
      "terminology": "26544005",
      "linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
      "previous_likelihood": 0.5
    },
    {
      "question": "Do your legs ever feel stiff or tight, making it hard to move as you wish?",
      "topic": "Motor Function",
      "target": "patient",
      "terminology": "7822000",
      "linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
      "previous_likelihood": 0.5
    },
    {
      "question": "Has your walking weakness worsened after periods of neck movement or certain postures?",
      "topic": "Motor Function",
      "target": "patient",
      "terminology": null,
      "linked_hypothesis": "Cervical spondylotic myelopathy",
      "previous_likelihood": 0.11
    },
    {
      "question": "Have you ever had sudden or severe pain radiating from your neck to your arms or back?",
      "topic": "Pain and Headaches",
      "target": "patient",
      "terminology": null,
      "linked_hypothesis": "Cervical spondylotic myelopathy",
      "previous_likelihood": 0.11
    },
    {

```



```

"question": "Have you had any changes in bladder or bowel function, such as urgency, frequency, retention, or
incontinence?",
"topic": "Autonomic Function",
"target": "patient",
"terminology": "300162001",
"linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
"previous_likelihood": 0.5
},
{
"question": "Have you recently had difficulty achieving or maintaining an erection or other changes in sexual
function?",
"topic": "Autonomic Function",
"target": "patient",
"terminology": "198053002",
"linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
"previous_likelihood": 0.5
}
]
}

```

A3.1:

```

{
"qa": [
{
"diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
"previous_likelihood": 0.5,
"question": "Have you noticed any changes in your mood, memory, or ability to think clearly over the past
several years?",
"answer": "Unknown",
"short_justification": "No data regarding cognitive or mood changes is provided.",
"evidence": null
},
{
"diagnosis": "Subacute combined degeneration (B12 or copper deficiency)",
"previous_likelihood": 0.2,
"question": "Have you experienced any episodes of confusion or unusual emotions recently?",
"answer": "Unknown",
"short_justification": "There is no information about confusion or emotional changes.",
"evidence": null
},
{
"diagnosis": "Cervical spondylotic myelopathy",
"previous_likelihood": 0.11,
"question": "Has anyone in your family had similar problems with movement or cognition?",
"answer": "Unknown",
"short_justification": "Family history is not provided.",
"evidence": null
},
{
"diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
"previous_likelihood": 0.5,
"question": "Have you ever had episodes of passing out, loss of consciousness, or seizure-like activity?",
"answer": "Unknown",
"short_justification": "No mention of seizure, syncope, or loss of consciousness.",
"evidence": null
},
{
"diagnosis": "Subacute combined degeneration (B12 or copper deficiency)",
"previous_likelihood": 0.2,
"question": "Have you ever lost consciousness suddenly or felt faint?",
"answer": "Unknown",
"short_justification": "There are no reported syncopal or fainting episodes.",
"evidence": null
},
{
"diagnosis": "Cervical spondylotic myelopathy",

```



```

"previous_likelihood": 0.11,
"question": "Have you had any episodes of sudden falls without warning?",
"answer": "Unknown",
"short_justification": "Falls without warning are not described.",
"evidence": null
},
{
  "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
  "previous_likelihood": 0.5,
  "question": "Do you experience any numbness, tingling, burning, or unusual sensations in your legs or hands?",
  "answer": "Yes",
  "short_justification": "The patient has paresthesias in hands and feet.",
  "evidence": "paresthesias in hands and feet (1 year before evaluation)"
},
{
  "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
  "previous_likelihood": 0.5,
  "question": "Have you noticed any recent difficulties feeling vibrations or knowing where your feet are?",
  "answer": "Yes",
  "short_justification": "Vibration and position sense are impaired in lower limbs.",
  "evidence": "Vibration sensation in the lower limbs was impaired in a graded manner up to the knees, and position sense was impaired at the toes and malleoli."
},
{
  "diagnosis": "Subacute combined degeneration (B12 or copper deficiency)",
  "previous_likelihood": 0.2,
  "question": "Have you experienced pins-and-needles sensations or loss of feeling in your hands or feet?",
  "answer": "Yes",
  "short_justification": "Paresthesias in hands and feet are documented.",
  "evidence": "paresthesias in hands and feet (1 year before evaluation)"
},
{
  "diagnosis": "Subacute combined degeneration (B12 or copper deficiency)",
  "previous_likelihood": 0.2,
  "question": "Have you observed any gradual loss of vibration or position sense in your toes or fingers?",
  "answer": "Yes",
  "short_justification": "Vibration and position sense impaired at toes and malleoli.",
  "evidence": "Vibration sensation in the lower limbs was impaired in a graded manner up to the knees, and position sense was impaired at the toes and malleoli."
},
{
  "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
  "previous_likelihood": 0.5,
  "question": "Have you felt muscle weakness, heaviness, or difficulty moving your legs?",
  "answer": "Yes",
  "short_justification": "Mild weakness in the lower limbs is observed.",
  "evidence": "mild weakness in the lower limbs"
},
{
  "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
  "previous_likelihood": 0.5,
  "question": "Do your legs ever feel stiff or tight, making it hard to move as you wish?",
  "answer": "Yes",
  "short_justification": "Increased tone and spasticity in lower limbs is described.",
  "evidence": "increased tone and mild weakness in the lower limbs, with an upper motor neuron pattern; spastic-ataxic gait"
},
{
  "diagnosis": "Subacute combined degeneration (B12 or copper deficiency)",
  "previous_likelihood": 0.2,
  "question": "Do you feel unsteady or do you lose your balance when standing or walking, especially in the dark?",
  "answer": "Yes",
  "short_justification": "The patient has a spastic-ataxic gait and a positive Romberg test.",
  "evidence": "He had a spastic-ataxic gait; The Romberg test was positive."
},
}

```



```
{
  "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
  "previous_likelihood": 0.5,
  "question": "Have you had trouble with your balance or fallen in the past year?",
  "answer": "Yes",
  "short_justification": "Spastic-ataxic gait and inability to walk tandemly suggest severe balance difficulties.",
  "evidence": "He had a spastic-ataxic gait and was unable to walk tandemly; The Romberg test was positive."
},
{
  "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
  "previous_likelihood": 0.5,
  "question": "Have you had any changes in bladder or bowel function, such as urgency, frequency, retention, or incontinence?",
  "answer": "Yes",
  "short_justification": "Urinary frequency and urgency are mentioned.",
  "evidence": "urinary frequency/urgency/hesitancy (1 year)"
},
{
  "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
  "previous_likelihood": 0.5,
  "question": "Have you recently had difficulty achieving or maintaining an erection or other changes in sexual function?",
  "answer": "Yes",
  "short_justification": "Erectile dysfunction is reported.",
  "evidence": "erectile dysfunction (5 years)"
}
]
```

A2-H5:

```
{
  "hypotheses": [
    {
      "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
      "rationale": "The patient's long-term high-dose zinc supplementation, progressive spastic-ataxic gait, lower limb upper motor neuron findings, sensory ataxia with severe vibration and position sense loss, bladder and erectile dysfunction, and distal paresthesias are consistently supported by multiple positive question-answer confirmations and detailed neurologic examination findings, making this the most likely diagnosis.",
      "likelihood": 0.8
    },
    {
      "diagnosis": "Subacute combined degeneration (B12 or copper deficiency)",
      "rationale": "Distal sensory loss, impaired proprioception and vibration sensation, spastic-ataxic gait, and a positive Romberg sign strongly support a myeloneuropathy due to B12 or copper deficiency. However, the strong association with chronic zinc supplementation favors copper deficiency secondary to zinc toxicity.",
      "likelihood": 0.35
    },
    {
      "diagnosis": "Cervical spondylotic myelopathy",
      "rationale": "Lower limb spasticity, mild weakness, impaired sensation, and progressive gait impairment are compatible with cervical myelopathy, but the absence of arm involvement and the compelling metabolic exposure history reduce its likelihood compared with nutritional etiologies.",
      "likelihood": 0.15
    }
  ]
}
```

A4:

```
{
  "investigations": [
    {
      "investigation": "Serum copper level",
      "batch_id": 1,
      "guideline_refs": "EAN/PNS 2021 CIDP Guidelines, AAN 2014 'Copper deficiency myelopathy', UpToDate 'Copper deficiency myeloneuropathy: Clinical manifestations, evaluation, and treatment', NICE NG145 1.4.1",
      "reason": "To confirm copper deficiency as the underlying cause of the myelopathy, likely secondary to chronic"
    }
  ]
}
```



```

excess zinc intake.",
  "urgency": "Urgent",
  "mandatory": "Mandatory",
  "timing": "Within 24–48 hours"
},
{
  "investigation": "Serum zinc level",
  "batch_id": 1,
  "guideline_refs": "AAN 2014 'Copper deficiency myelopathy', UpToDate 'Copper deficiency myeloneuropathy: Clinical manifestations, evaluation, and treatment'",
  "reason": "To document zinc toxicity, supporting the suspected diagnosis of zinc-induced copper deficiency myeloneuropathy.",
  "urgency": "Urgent",
  "mandatory": "Mandatory",
  "timing": "Within 24–48 hours"
},
{
  "investigation": "Serum vitamin B12 level",
  "batch_id": 2,
  "guideline_refs": "NICE CG153, UpToDate 'Diagnosis of vitamin B12 and folate deficiency', AAN 2014 'Copper deficiency myelopathy'",
  "reason": "To rule out or identify vitamin B12 deficiency as a cause of subacute combined degeneration, given overlapping clinical features.",
  "urgency": "Routine",
  "mandatory": "Mandatory",
  "timing": "Within 1 week"
},
{
  "investigation": "Serum methylmalonic acid and homocysteine",
  "batch_id": 2,
  "guideline_refs": "UpToDate 'Diagnosis of vitamin B12 and folate deficiency', NICE CG153",
  "reason": "To confirm or exclude functional B12 deficiency, especially if serum B12 is borderline or near the lower reference range.",
  "urgency": "Routine",
  "mandatory": "Optional",
  "timing": "Within 1 week"
},
{
  "investigation": "MRI cervical spine with and without contrast",
  "batch_id": 3,
  "guideline_refs": "EAN 2022 Consensus 'Diagnosis of non-compressive myelopathies', AAN 2014 'Evaluation of suspected myelopathy'",
  "reason": "To evaluate for alternative or additional causes of myelopathy, including cervical spondylotic myelopathy, demyelination, or posterior column signal changes typical of copper or B12 deficiency.",
  "urgency": "Routine",
  "mandatory": "Mandatory",
  "timing": "Within 2 weeks"
},
{
  "investigation": "Neurophysiology: nerve conduction studies (NCS) and electromyography (EMG)",
  "batch_id": 4,
  "guideline_refs": "EAN/PNS 2021 CIDP Guidelines, AAN 2014 'Evaluation of suspected myelopathy'",
  "reason": "To assess for associated peripheral neuropathy and to characterize the pattern of nerve involvement (axonal versus demyelinating).",
  "urgency": "Routine",
  "mandatory": "Mandatory",
  "timing": "Within 4 weeks"
}
]
}

A5:
{
  "treatments": [
    {
      "treatment": "Cessation of zinc supplementation",

```



```

    "dose": "Discontinue immediately and permanently",
    "timing": "Immediate",
    "reason": "High-dose zinc interferes with copper absorption and induces copper deficiency; immediate
cessation is foundational and universally recommended.",
    "linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
    "mandatory": "Mandatory"
  },
  {
    "treatment": "Copper supplementation (oral or intravenous)",
    "dose": "Oral copper gluconate or sulfate providing 2 mg elemental copper daily; if severe deficiency or
impaired gastrointestinal absorption, 2–4 mg IV copper daily for several days until serum copper normalizes,
followed by oral maintenance at 2 mg daily",
    "timing": "Initiate as soon as copper deficiency is confirmed; consider empiric supplementation if clinical
suspicion is high and no contraindications exist",
    "reason": "Copper replacement corrects hematologic abnormalities and may stabilize or partially reverse
neurological deficits; earlier initiation is associated with improved neurological outcomes.",
    "linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
    "mandatory": "Mandatory"
  },
  {
    "treatment": "Physical therapy and gait rehabilitation",
    "dose": "N/A",
    "timing": "Initiate after acute management and stabilization, typically within 2–4 weeks",
    "reason": "To optimize functional recovery, improve strength and balance, reduce spasticity, and address gait
ataxia following correction of the underlying deficiency.",
    "linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
    "mandatory": "Mandatory"
  },
  {
    "treatment": "Watchful waiting with structured follow-up",
    "dose": "N/A",
    "timing": "Ongoing monitoring over weeks to months after correction of deficiency",
    "reason": "Neurological recovery may be partial and delayed; close monitoring allows assessment of response
to therapy and guides further rehabilitation or diagnostic reassessment if progression occurs.",
    "linked_hypothesis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
    "mandatory": "Mandatory"
  }
]
}

```

A5.5:

```

{
  "safety_check": {
    "drug_interactions": [
      {
        "drug1": "zinc",
        "drug2": "copper",
        "severity": "major",
        "mechanism": "Zinc competitively inhibits copper absorption in the gastrointestinal tract, leading to copper
deficiency.",
        "clinical_effect": "Concurrent zinc use reduces the effectiveness of copper supplementation and risks
persistent or worsening copper deficiency.",
        "management": "avoid"
      }
    ],
    "contraindications": [
      {
        "drug": "zinc",
        "reason": "Zinc is absolutely contraindicated in proven or suspected copper deficiency due to the risk of
worsening neurologic and hematologic manifestations.",
        "type": "absolute",
        "alternative_options": [
          "Cessation of zinc supplementation"
        ]
      }
    ]
  }
}

```



```

"dose_adjustments": [
  {
    "drug": "copper",
    "factor": "age",
    "original_dose": "2 mg elemental copper daily (oral), or 2–4 mg IV for severe cases",
    "adjusted_dose": "Initiate at 2 mg elemental copper daily; consider starting at the lower end of the dose range and titrate based on laboratory values and tolerability",
    "rationale": "Older adults may have altered pharmacokinetics and increased susceptibility to adverse effects, warranting cautious initiation and close monitoring.",
    "monitoring_required": [
      "serum copper",
      "ceruloplasmin",
      "CBC",
      "LFTs"
    ]
  }
],
"therapeutic_duplications": [],
"monitoring_plan": [
  {
    "parameter": "serum copper",
    "frequency": "Baseline; weekly until normalization, then monthly",
    "target_range": "70–140 mcg/dL"
  },
  {
    "parameter": "ceruloplasmin",
    "frequency": "Baseline; repeat every 2–4 weeks until stable",
    "target_range": "17–54 mg/dL"
  },
  {
    "parameter": "CBC",
    "frequency": "Baseline; every 2 weeks during replacement, then every 2–3 months",
    "target_range": "Hemoglobin, WBC, and platelets within age-appropriate reference ranges"
  },
  {
    "parameter": "LFTs",
    "frequency": "Baseline; then monthly during copper supplementation",
    "target_range": "Within normal limits for age"
  },
  {
    "parameter": "Neurological examination",
    "frequency": "Every 4–8 weeks for the first 6 months, then as clinically indicated",
    "target_range": "Stabilization or improvement of neurological deficits"
  }
],
"safe_to_proceed": true,
"requires_modification": false,
"pharmacy_consultation_needed": false,
"high_alert_medications": [
  "copper (IV or high-dose oral)"
]
}

```

A6:

```

{
  "validation_report": {
    "is_complete": true,
    "completeness_score": 1.0,
    "missing_elements": [],
    "inconsistencies": [],
    "critical_diagnoses_check": {
      "all_life_threatening_considered": true,
      "missed_critical": []
    }
  },
  "guideline_compliance": {

```



```

"guidelines_cited": 6,
"compliance_rate": 1.0,
"missing_citations": []
},
"safety_net_included": true,
"patient_instructions_clear": true,
"follow_up_specified": true,
"confidence_score": 1.0,
"requires_human_review": false,
"review_urgency": "routine",
"quality_flags": []
}
}

```

DiagnosticTestExtractor:

```

{
  "thoughts": [],
  "thoughts_internal": [],
  "summary": {
    "message": "No additional diagnostic test results were provided in these sections."
  }
}

```

HypothesisRefiner:

```

{
  "summary": {
    "message": "No additional diagnostic tests or clinical information were received for further refinement.",
    "hypotheses": {
      "hypotheses": [
        {
          "diagnosis": "Zinc-induced myelopathy (copper deficiency myeloneuropathy)",
          "rationale": "The patient's long-term high-dose zinc supplementation, progressive spastic-ataxic gait, lower limb upper motor neuron findings, sensory ataxia with severe vibration and position sense loss, bladder and erectile dysfunction, and distal paresthesias are all directly supported by multiple positive QA responses and the provided neurologic examination evidence, making this the most likely diagnosis.",
          "likelihood": 0.8
        },
        {
          "diagnosis": "Subacute combined degeneration (B12 or copper deficiency)",
          "rationale": "Distal sensory loss, impaired proprioception and vibration sensation, spastic-ataxic gait, and a positive Romberg sign strongly support a myeloneuropathy due to B12 or copper deficiency. However, the strong association with chronic zinc supplementation favors copper deficiency secondary to zinc toxicity.",
          "likelihood": 0.35
        },
        {
          "diagnosis": "Cervical spondylotic myelopathy",
          "rationale": "Lower limb spasticity, mild weakness, impaired sensation, and progressive gait impairment are compatible with cervical myelopathy, but the absence of arm involvement and the compelling metabolic exposure history reduce its likelihood compared with nutritional etiologies.",
          "likelihood": 0.15
        }
      ]
    }
  }
}

```



Validation Platform for Agent-Specific Mapping Accuracy

To systematically evaluate whether agent-specific prompts correctly map free-text clinical narratives into structured representations, we developed a dedicated agent output validation platform that is already deployed⁴, but also available as an open-source code at the ChatMED project GitHub repository⁵. This platform serves as an operational validation layer for assessing the fidelity, completeness, and controllability of the mapping process performed by each agent in the Neurology Orchestrator workflow.

The core objective of the platform is to verify that each agent:

1. Extracts only information explicitly present in the source narrative,
2. Maps that information exclusively to its predefined semantic scope, and
3. Produces a structured JSON output aligned with its functional specification, without hallucinating, inferring unsupported facts, or introducing extraneous content (“noise”).

Rather than evaluating downstream clinical correctness, the platform focuses specifically on prompt–agent alignment, i.e., whether the prompt defining an agent’s role enables a controlled and semantically faithful transformation from free-text input to structured data. Accordingly, this validation step ensures that datasets used for subsequent fine-tuning of the NeuroOrchestrator are generated under controlled conditions and are rigorously validated before further use.

Platform Overview

The verification platform provides a centralized interface for:

- importing or loading clinical cases,
- tracking validation status across multiple agents,
- inspecting agent-specific outputs in isolation, and
- performing structured human validation at scale.

At the global level, the platform maintains a **case-centric dashboard** (Figure 4) summarizing:

- total number of evaluated cases,
- pending, partially validated, and fully validated cases,
- per-case validation progress across agents.

⁴ <https://validation.chatmed-project.eu>

⁵ https://github.com/ChatMED/validation_platform

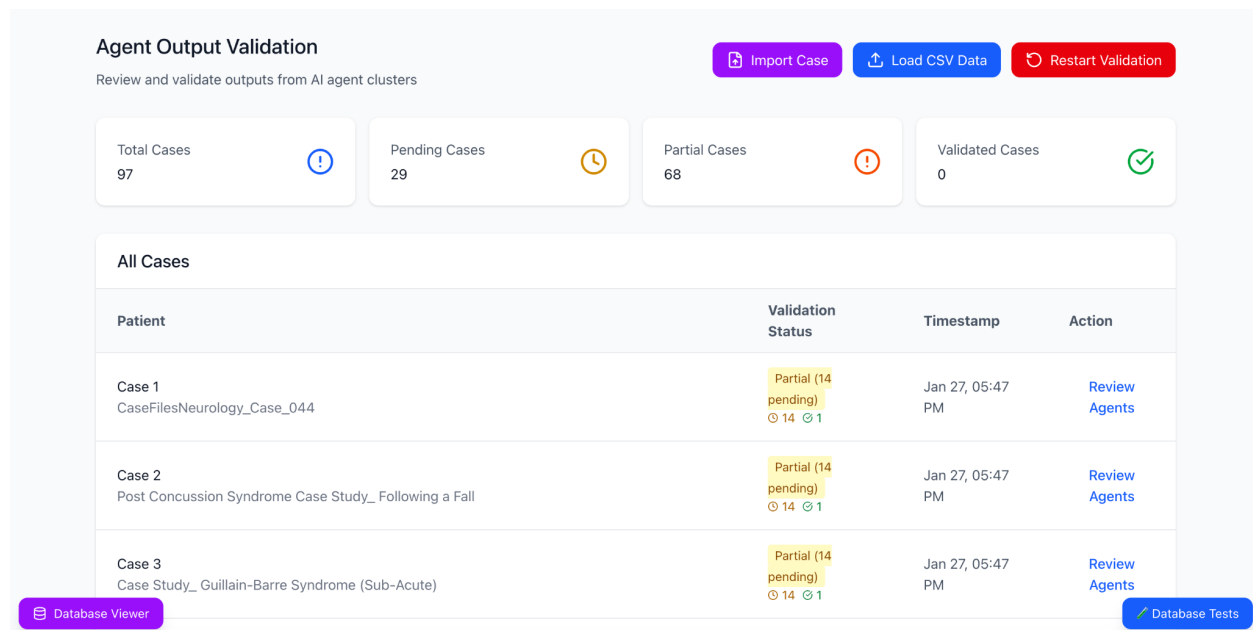


Figure 4. The case-centric dashboard.

Case-Level Validation Workflow

The platform enables both macro-level monitoring (e.g., validation throughput, bottlenecks) and micro-level inspection of individual agent outputs.

Each clinical case is represented by **(Figure 5)**:

- a reference diagnosis (used only for contextual grounding, not for agent evaluation),
- a difficulty annotation (e.g., basic, intermediate, complex),
- the original free-text clinical narrative, preserved verbatim.

The narrative acts as the single source of truth against which all agent outputs are evaluated.

For each case, the platform decomposes validation into agent-specific units, corresponding to the functional agents in the Neurology Orchestrator (e.g., metadata extraction, risk assessment, anatomical localization, pathophysiology modeling, differential diagnosis generation, diagnosis ranking).

Each agent produces:

- a structured JSON file,
- generated exclusively from the free-text narrative using its dedicated prompt.

No agent has access to other agents' outputs during generation, ensuring independence and preventing leakage or cross-contamination.

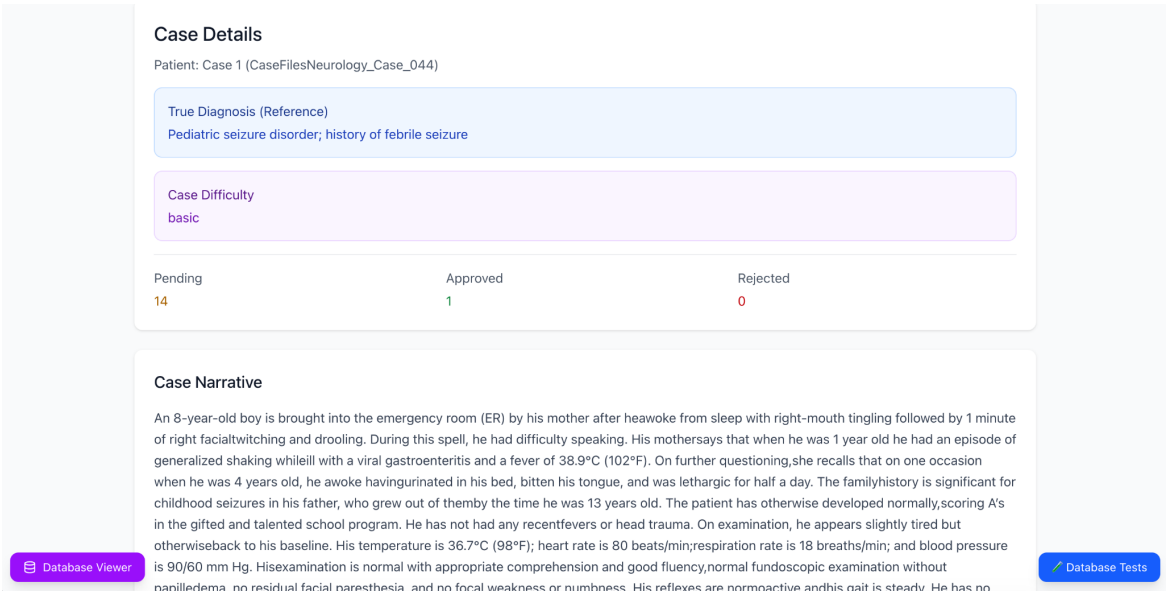


Figure 5. Case-level view.

Agent Output Inspection and Validation

For every agent (Figure 6), the platform exposes (Figure 7, Figure 8, Figure 9):

- the generated JSON output,
- the agent’s functional role,
- the prompt-defined task scope,
- and the current validation status (pending, approved, rejected).

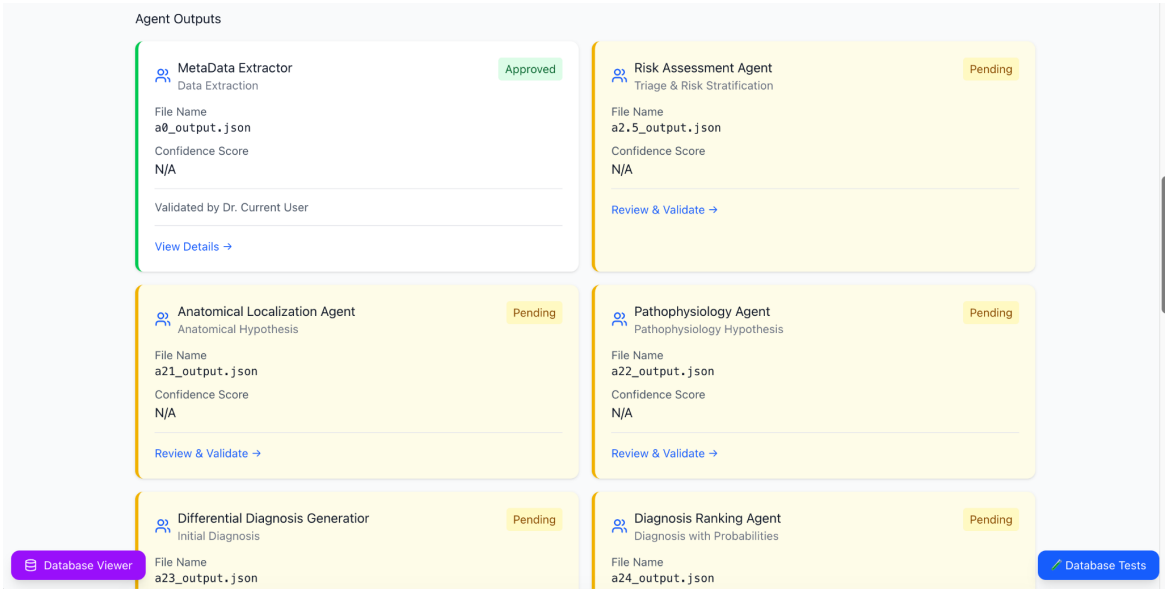


Figure 6. Agent-level view.

Validation is performed by human reviewers through the action **Review & Validate**, upon which the outputs are assessed by using a free-text narrative of a doctor.

Validation decisions are recorded per agent per case, enabling fine-grained auditing. Partial validation states naturally emerge when some agents are approved while others remain under review.

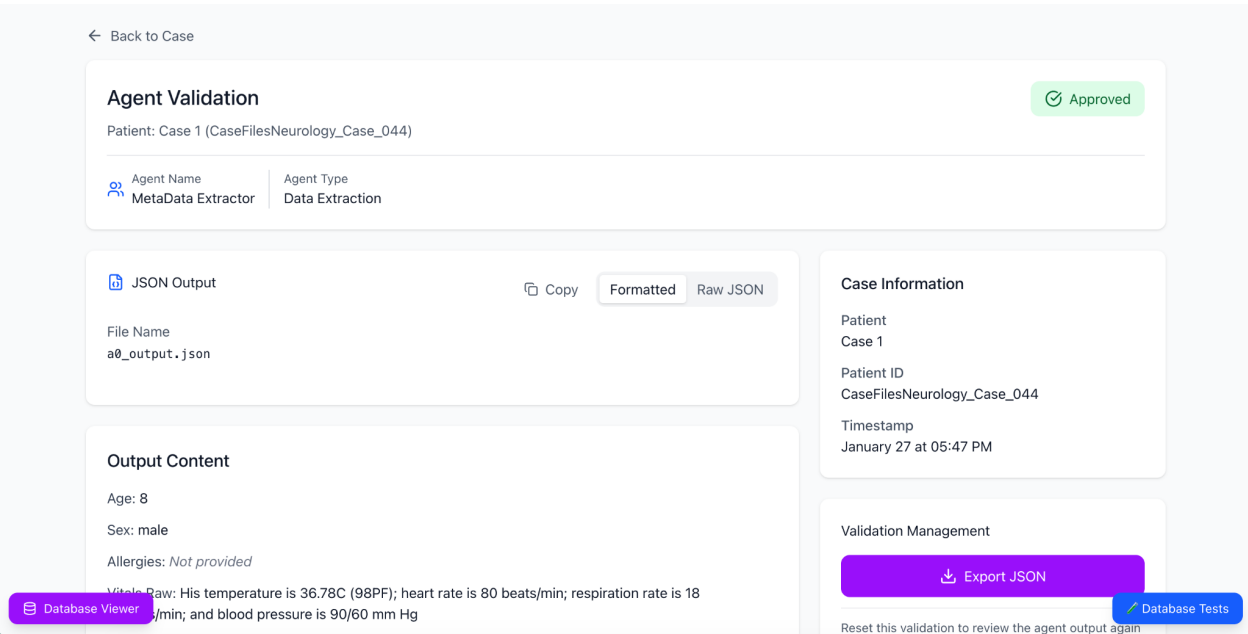


Figure 7. Agent validation example (a).

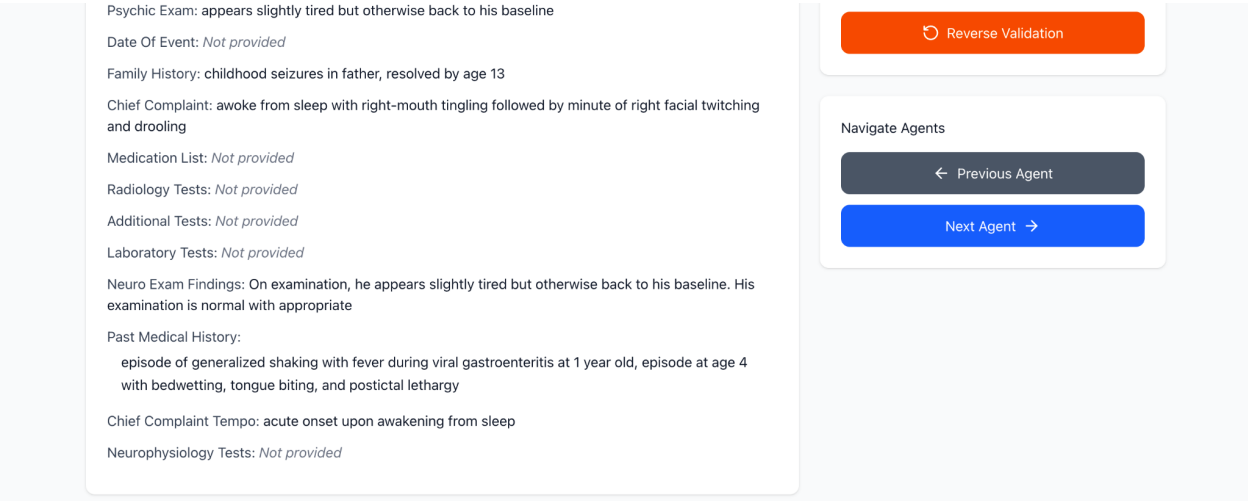


Figure 8. Agent validation example (b).

Past Medical History:

episode of generalized shaking with fever during viral gastroenteritis at 1 year old, episode at age 4 with bedwetting, tongue biting, and postictal lethargy

Chief Complaint Tempo: acute onset upon awakening from sleep

Neurophysiology Tests: *Not provided*

Validation Notes Included in export

These notes will be included in the exported JSON file along with validation metadata. You can continue to edit notes even after validation.

For NeuroExam finding output get only first sentence which is general observation and sad "examination is normal with appropriate". That's generally true summary for sentences "His examination is normal with appropriate comprehension and good fluency, normal fundoscopic examination without papilledema, no residual facial paresthesia, and no focal weakness or numbness. His reflexes are normoactive and his gait is steady. He has no skin lesions. He has no evidence of..."

 Save Notes

Validated by Dr. Current User on January 29, 2026 at 06:37 PM

 Database Viewer  Database Tests

Figure 9. Agent validation example (c).

Role of the Platform in the Methodology Refinement

The platform enabled early identification of:

- prompts that consistently over-extract or hallucinate,
- agents that require tighter instruction constraints,
- and cases where narrative ambiguity challenges structured extraction.

Importantly, the platform is not merely an evaluation tool but a **methodological feedback mechanism**. Validation outcomes are used to:

- iteratively refine agent prompts,
- adjust JSON schemas for clarity and minimalism,
- and recalibrate agent boundaries to reduce overlap or ambiguity.

Through this loop, prompt design, agent specification, and validation are tightly coupled, ensuring that improvements are evidence-driven rather than ad hoc. This approach is of essential importance for reliability confirmation of the dataset created in D7.2.



Conclusion

This deliverable consolidates the methodological groundwork for the NeuroOrchestrator as a structured, neurology-driven reasoning pipeline designed to transform free-text clinical narratives into transparent, modular, and verifiable decision-support steps. Rather than treating neurological cases as unstructured text-to-answer problems, D7.1 formalizes expert neurological reasoning as a sequence of explicitly defined agentic steps (A0–A7), each with a clear clinical role, constrained prompt specification, and machine-readable output schema.

A central contribution of this deliverable is the demonstration that complex neurological reasoning can be decomposed into interpretable intermediate representations without sacrificing clinical fidelity. The detailed scenario mapping illustrates how narrative-rich cases from established neurology literature can be reliably translated into structured reasoning traces that preserve clinical nuance while enabling computational transparency and auditability.

Equally important, D7.1 establishes a rigorous validation layer through the introduction of a dedicated verification platform for agent-specific validation. By focusing on prompt–agent alignment rather than downstream diagnostic correctness, the platform ensures that each agent’s output remains faithful to the source narrative, adheres strictly to its semantic scope, and avoids hallucination or uncontrolled inference. This human-in-the-loop validation mechanism closes a critical gap between prompt engineering and dataset construction, providing controlled, reviewable, and reproducible reasoning artifacts.

Through this combined methodology and validation approach, the deliverable defines a reliable pathway from raw clinical text to training-grade datasets. The NeuroOrchestrator schema, prompt definitions, JSON output structures, and validation procedures together constitute a coherent data-generation framework suitable for fine-tuning large language models toward expert-level neurological reasoning. Importantly, this pipeline supports scalability: from minimal vignette-style inputs to full narrative cases, and from pilot-scale validation to larger curated datasets.

In this sense, D7.1 does not aim to deliver a finalized dataset, but rather to establish a robust methodology for the representation of neurological scenarios, which will subsequently be applied to map hundreds/ thousands of neurological cases into a reasoning-based dataset. This approach ensures that downstream model training is grounded in clinically validated reasoning traces, rather than in opaque or weakly supervised text correlations. It therefore positions the consortium to proceed confidently toward Deliverable D7.2, where dataset expansion and annotation consistency will be formalized, and toward WP8, in which the NeuroOrchestrator will be instantiated as a fine-tuned, reasoning-aware component within the broader ChatMED multi-agent ecosystem.