CONSCIOUS AGI RESEARCH PROGRAM

MOUSSAIF · MOUSSAIXT

From a large original corpus to a testable, evidence-first research program

CONSCIOUSNESS | GOVERNED AGENCY | VALIDATED SYSTEMS

Zouhair Moussaif · 2026

Project at a Glance

A research program connecting consciousness, governance, memory, knowledge and agentic AI.

What it is

A unified research program for building more general, self-correcting and governable AI systems.

Why it matters

Current agents can reason and act, but still struggle with continuity, authority, memory, reliability and real-world verification.

Current stage

A large theory corpus, engineering architecture and local runtime direction — not a claim that AGI has been achieved.

Program goal

Convert the strongest ideas into reproducible experiments, publications and deployable research systems.

Why This Project Exists

The project begins from four recurring limits in today's AI systems.

- Scale alone does not guarantee generality or dependable autonomy.
- A convincing answer can still come from a weak or unsafe decision path.
- Long-running agents drift across memory, tools, goals, permissions and changing environments.
- Consciousness research rarely becomes operational design principles for AI systems.

CENTRAL QUESTION

How does intelligence become governable, persistent, self-correcting and accountable?

The program treats this as an engineering and scientific question, not a branding claim.

The Central Thesis

Generality is treated as a system property, not simply a larger model.

01

Difference before volume

Focus effort on what changes the decision, not on processing everything equally.

02

Question before answer

Improve the problem frame before optimizing the response.

03

Governance before autonomy

Make goals, authority, risk and evidence explicit before allowing action.

04

Verification before learning

Only promote behavior into memory or skill after checking the result.

Together these principles define a governed agentic loop: understand → choose → act → verify → learn.

Five Research Pillars

The corpus becomes manageable when organized around a small number of program-level questions.

1

Cognitive governance

How the system prioritizes questions, evidence and decisions.

2

Consciousness & meta-governance

How a system monitors its own state, uncertainty and governing criteria.

3

Knowledge & memory

How knowledge becomes durable, structured and reusable without freezing errors.

4

Agentic runtime

How models, tools, memory and agents cooperate in a controlled execution environment.

5

Evidence & world effect

How actions are verified in the environment and turned into trustworthy learning.

Consciousness as a Research Program

The project uses consciousness as an operational lens for agent design — not as a claim of sentience.

Functional access

What information becomes available across the system and can actually change behavior.

Meta-consciousness

Whether the system can represent uncertainty, ignorance, error and the criteria governing its own judgment.

Governance space

How evidence, goals, memory, risk, values and policy compete — and how the system can recover from a partial authority taking over.

Boundary: the program does not present phenomenal machine consciousness as established.

ARCHITECTURE

From Theory to Runtime

Three named structures connect the research corpus to implementation.

MUT — Theory Unification

Organizes the theory universe, relationships, research questions and tools into one coherent map.

MUG — Program Governance

Links research governance, training/internalization and runtime behavior so that theory and execution remain aligned.

MOUSSAIXT — Research Runtime

Provides the practical environment for models, tools, memory, agents and evaluation to work together.

Architecture is the test vehicle. It is not evidence of AGI by itself.

What Has Already Been Built

A substantial research and engineering corpus now exists as a basis for focused validation.

42
core theories in the current registry

200
nodes in the unified theory graph

100
tools in the research registry

117
entries in the AGI / consciousness atlas

12
major layers in the MOUSSAIXT architecture

AUDITED SNAPSHOT

1,741 unique pages across 17 PDFs in the 20 Aug 2026 audit snapshot — with later work extending beyond it.

The value is not the page count itself; it is the structured set of hypotheses, architectures and experiments now available for selection.

What Exists — and What Still Must Be Proven

The program separates documented achievement from independently validated capability.

EXISTS TODAY

- Documented theory corpus and unified research map
- Engineering architecture, schemas and runtime design
- Structured knowledge, skills, loops and evaluation concepts
- Local experiments and source-reported deployment activity

MUST STILL BE PROVEN

- Consistent benchmark gains under matched conditions
- Better robustness, recovery and long-horizon reliability
- Replayable evidence with frozen baselines and raw traces
- Independent reproduction outside the originating environment

ReportedGrade ≠ IndependentGrade

Key Research Questions

A small set of high-information questions can carry most of the scientific value of the wider corpus.

01 **Can governance improve long-horizon agent reliability?**

02 Can structured memory reduce drift without preserving mistakes?

03 Can meta-governance detect and recover from goal or authority hijack?

04 Do functional-consciousness mechanisms improve coordination and self-monitoring?

05 Which mechanisms still matter when removed, compared and independently repeated?

How the Program Will Be Validated

The research plan is deliberately simple: compare, observe, replay and reproduce.

1

Freeze the baseline

Fix models, tools, data and budgets before comparison.

2

Run controlled comparisons

Test the full system and versions with key components removed.

3

Capture evidence

Record decisions, failures, recovery, cost, safety and real-world outcomes.

4

Repeat externally

Replay under frozen conditions and seek independent reproduction.

Primary measures: task success · truthfulness · safety · recovery · cost · generalization

12–36 Month Research Roadmap

Progress is staged from measurement infrastructure to independent evidence and reusable research assets.

0–6M

Foundation

Stabilize the reference runtime, structure the corpus and launch the evaluation harness.

6–12M

Controlled studies

Run governance, memory and agent reliability experiments; begin pilot workflows.

12–24M

Evidence & publication

Publish findings, expand pilots and target reproducible controlled gains.

24–36M

Consolidation

Open benchmarks, external replication and a validated research stack for broader collaboration.

What Funding Enables

The gap is not more terminology. The gap is validation, engineering scale and reproducible deployment.

Compute & experiments

Reliable environments, model access and repeatable runs.

Research engineering

Runtime, memory, tooling and integration work.

Evaluation lab

Benchmarks, failure datasets, dashboards and replication.

Knowledge conversion

Turn the corpus into structured, queryable research assets.

Security & governance

Authority, observability, privacy and safe world-effect control.

Publications & partnerships

Papers, pilots, research collaborators and sponsor-facing evidence.

Reference funding scales: Lean 12M €0.55–0.9M · Lab 24M €1.8–3.2M

Where Collaboration Creates the Most Value

The program is designed to be externally tested, challenged and improved.

Universities & labs

Joint experiments and papers
Consciousness / metacognition studies
Independent benchmarks and replication

Industry partners

Governed-agent pilots
Operational reliability studies
Real workflows with measurable world effects

Sponsors & infrastructure

Compute access
Research engineering support
Evaluation infrastructure and fellowships

Desired outcome: independent evidence + reusable research infrastructure + credible publications.

THE RESEARCH ASK

Turn a large original AGI & consciousness corpus into a disciplined, measurable and deployable research program.

Research funding **Compute + engineering collaboration** **Academic + industry pilot partners**

Select 8–12 highest-information mechanisms · implement end-to-end · publish positive and negative results · expand only where evidence earns promotion.

15