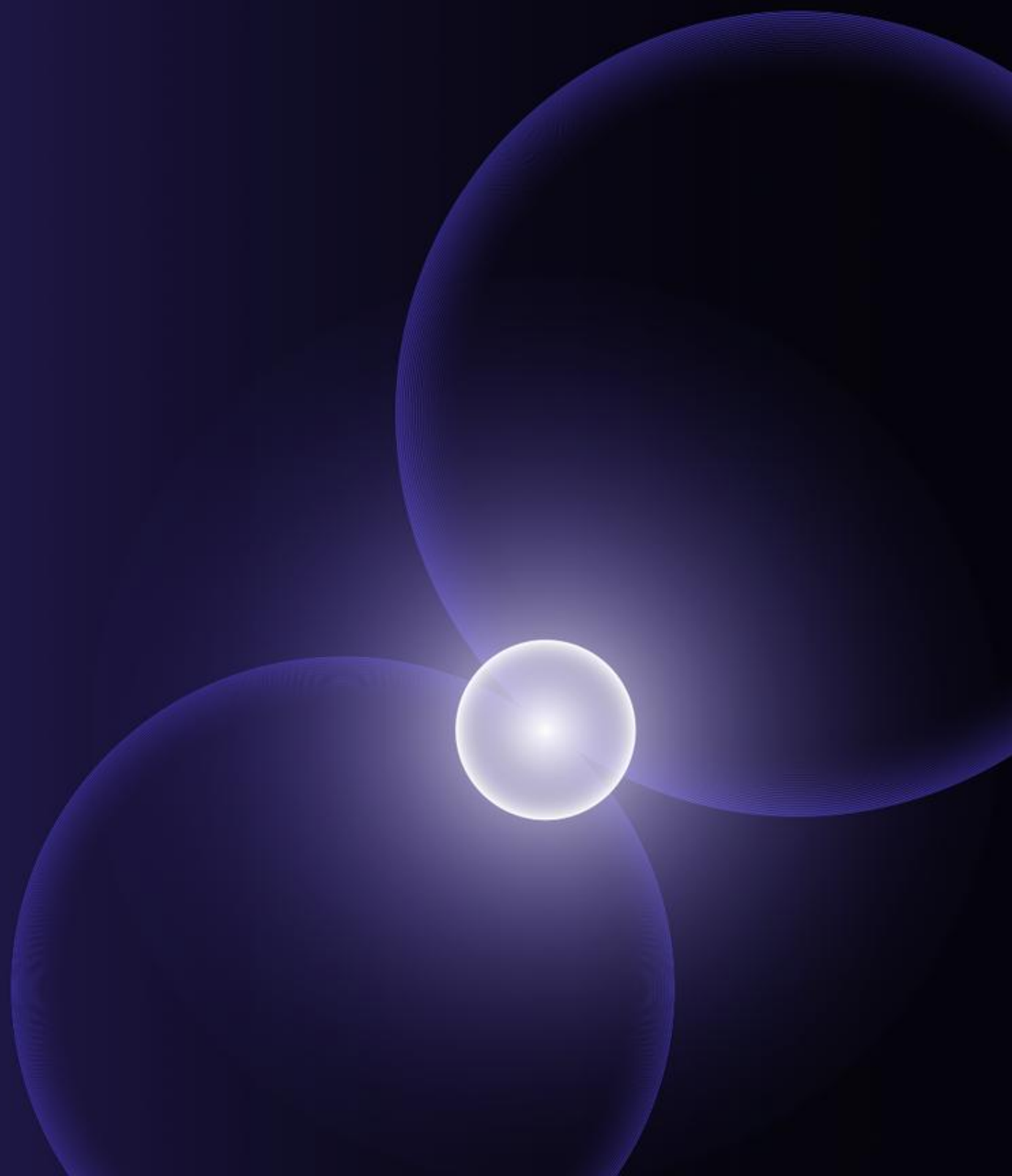
conscium

Latin: /ˈkɒn.ʃum/

Adjective: conscious, aware, knowing

The Future of AI Verification

Pioneering Safe,
Efficient AI



Agenda

- Intro to Conscium
- Simulation-based Verification
- Our principles for responsible AI consciousness R&D

Founding Team

Dr. Daniel Hulme

Founded and sold AI consultancy Satalia to WPP in 2021 for $100m. Currently Chief AI Officer at WPP. A pioneer in Artificial Life.

Ass. Prof. Ted Lappas

An expert in neural architectures for multi-modal data. Currently leads data science team at WPP. Ted is recognised by Stanford University in the top 2% of scientist in his field in both 2023 and 2024.

Ass. Prof. Panos Repoussis

An expert in evolutionary computation and data-driven optimisation. Currently leads AI Research Lab at WPP.

Ed Charvet

A serial entrepreneur with 2 exits, the last to Datatec, where he became CSO and then COO EMEA for Logicalis, its multi-billion dollar tech subsidiary. He is an angel and Chair of the VC backed med tech platform, PreActiv.

Calum Chace

Previously CEO of a strategy consulting firm, Calum is a keynote speaker, adviser, and best-selling writer on artificial intelligence. He has is a recognised global speaker on AI with over 150 talks in 20 countries. He co-hosts the London Futurists Podcast. Calum studied PPE at Oxford.

WPP

Co-founder, investor and shareholder. Providing access to clients and distribution channels, management talent and expertise for Conscium

Problems with current AIs

Passive Training

Current AIs are trained passively on historical data, which makes it difficult to eliminate the bias in the data, and avoid IP infringement.

Seeking Sustainability

Today's leading AI models demand massive volumes of training data, huge compute power and energy.

No Focus on Consciousness

The failure to investigate machine consciousness could allow mind crime to occur on a massive scale, and could neglect an important route to safer superintelligence.

Static Neural Architectures

Current AI architectures are rigid, and unable to adapt their neural connections to develop new capabilities. These limitations persist despite increases in compute and training data.

Lack of Transparency

Current AIs are complex black-box models. We do not understand how they make decisions or what they are capable of achieving.

Conscium’s north star and mission

Understanding the nature of consciousness is among humanity's most important quests. What is consciousness? Can we create conscious AI? The answers to these questions have a material impact on who we are as a species, and our place in the universe.

Conscium is dedicated to deepening our understanding of artificial consciousness - its feasibility and implications. In doing so, we will create a profitable business, based on agent verification and advances in neuromorphic computing.

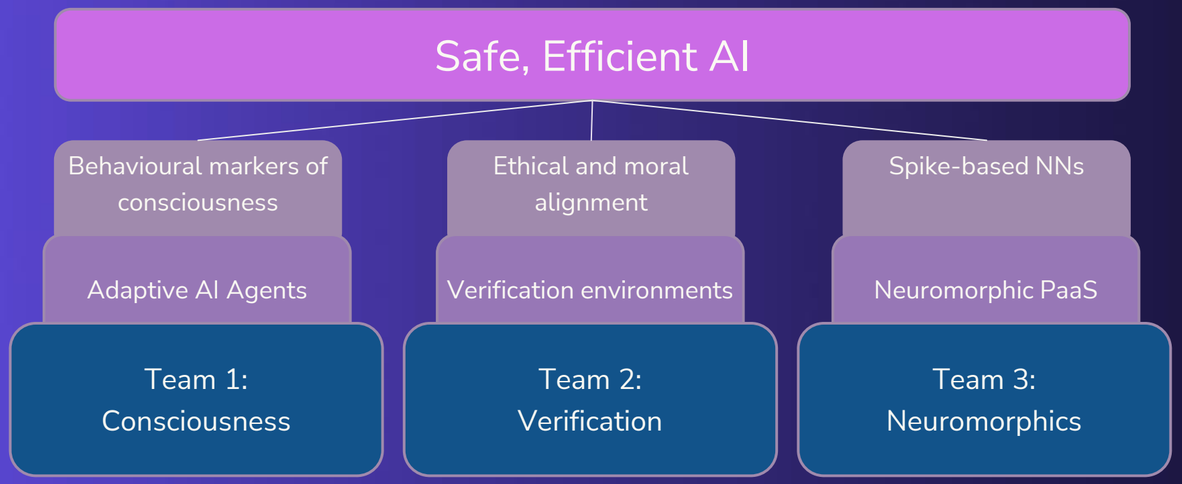
The importance of verification

The age of agentic AI has begun. These agents, capable of complex tasks, are permeating every aspect of our lives.

Given their autonomy and their impact, it is vital to verify that they do what they are designed to do, safely.

Through WPP, Conscium has a vast platform to develop a wide range of verification scenarios, starting with marketing and communication services. It also has a channel to reach the largest users of AI agents today.

Summary of Conscium’s Proposition



Conscium is in the business of providing safe, efficient AI by combining the disciplines of neuromorphics, verification and consciousness research

Conscium will deliver a platform for the development and verification of efficient, adaptive and aligned AI agents. The platform include modular offerings around verification and agent development to capture a board market

These agents will power future applications across many sectors - including transport, robotics, intelligent sensing, medical solutions

Conscium's Journey Has Begun

Solving Machine Consciousness

Produced white paper entitled "Solving Machine Consciousness", assessing approaches and describes our road map

Principles for AI Consciousness Research

Produced white paper on principles for approaching AI consciousness research with Dr Patrick Butiln of Oxford University

Advisory Board Recruited

Recruited some of the world's leading neuroscientists and computer scientists to our advisory board

Working with Prof Mark Solms & his Lab

Prof Mark Solms, his lab and work on self sufficient adaptive AI agents is the kernel of Conscium's Consciousness Lab
(Team 1)

Developing ethics platform for AI alignment

Developed "Moral Me", a platform to gather behavioural and values data to enable AI alignment within verification environments

Verification Platform is currently underway. Expecting to verify WPP knowledge-base agents by end of June 2025
(Team 2)

Demonstrations of Neuromorphic capabilities

Working closely and co-maturing projects with key academics and their teams (Prof Nikola Kasabov, Prof. Steve Furber & Michael Hopkins, Prof. Jason Eshragian)
(Team 3)

Advisory Board

Prof Karl Friston
Neuroscientist

https://en.wikipedia.org/wiki/Karl_J._Friston

Prof Mark Solms
Neuropsychologist

https://en.wikipedia.org/wiki/Mark_Solms

Prof Steve Furber
Neuromorphic Computing Scientist

https://en.wikipedia.org/wiki/Steve_Furber

Prof Nick Humphrey
Philosopher and Neuroscientist

https://en.wikipedia.org/wiki/Nicholas_Humphrey

Prof Benji Rosaman
Computer Scientist

https://www.wits.ac.za/people/academic-a-z-listing/r/benjaminrosman1witsacza/

Prof Nikola Kasabov
Knowledge Engineer

https://en.wikipedia.org/wiki/Nikola_Kasabov

Prof Nicola Clayton
Cognitive Scientist

https://en.wikipedia.org/wiki/Nicky_Clayton

Dr Suzanne Livingston
Philosopher and Neuroscientist

https://www.drsuzannelivingston.com/

Ass. Prof Megan Peters
Cognitive Scientist

https://faculty.uci.edu/profile/?facultyId=6594

Ass. Prof Jason Eshraghian
SNNTorch founder

https://ncg.ucsc.edu/jason-eshraghian-bio/

Prof Moran Cerf
Neuroscientist

https://en.wikipedia.org/wiki/Moran_Cerf

Sir Anthony Finkelstein
Former Chief Scientific Advisor UK Gov.

https://en.wikipedia.org/wiki/Anthony_Finkelstein

Prof Anil Seth
Neuroscientist

https://en.wikipedia.org/wiki/Anil_Seth

Ass. Prof Jonathan Shock
Applied Mathematician

https://shocklab.net/

Verification Levels

Level 1: Verify knowledge and perform basic tests (consistency, robustness etc)

Level 2: Tools usage & multiple tool workflows

Level 3: Reasoning and planning on real-world scenarios

Level 4: Internal states (morality, ethics, etc)

Level 5: Consciousness

SImulation-Based Verification

What is a simulation environment?

- Imagine a simulation game with an arbitrary number of autonomous agents.
  - 1 or more black-box Main Character (MC) / Client agents
  - Non-Player Characters (NPC) virtual agents

- Each NPC has its own clock, a personality, objective(s), a memory (stores selectively the events, observations and info collected), a list of resources, and a list of agents that it can interact with (neighbors). It can interact with an agent or resource at any point (like independent sensors in a distributed network)
  - Resources can be data sources or tools (APIs).
  - Access to other agents may be "read" or "write".

- Let neighboring agents X and Y:
  - Agents may communicate (SPEAK) via text, audio, images or video.
  - Agent Y can LISTEN to X (=receive messages from it) OR SPEAK to X (=send messages to it) OR both.
  - Agent Y if it has “write” rights it can ALTER X personality, objective, access to resources/neighbors, etc.

Simulation-Based Verification

The Game Maker Agent (can be also a human domain expert)

- Creates the game
  - Initiates all NPCs based on a predefined character library and defines the universe of interactable resources (data & tools)
  - Initializes the memory of each agent by communicating some info ("context") and objectives
  - Generates the ground truth, benchmarks and defines difficulty / complexity

- Does not interact with other agents during the simulation. It creates everything and then switches off.

- Sets the global simulation clock (start and end of the simulation). The agents are not aware of this global clock. On the global clock signals that the simulation is over, the all agents switch off.

Simulation-Based Verification

The Evaluator Agent (post-simulation)

- Has access to all available information / attributes related to agents, their memory and the resources.

- Does not interact with any agents, but it has access to the benchmarks produced by the Game Maker.

- Studies all available info from the NPCs + testimonies from the MC agents and creates a report.
  - It compiles the verification report and generates metrics, visualizations and analytics.
  - Post-simulation interview session with MC agents to get their point of view and enrich its data.

Simulation-Based Verification

The God/Storyteller Agent (can be seen as an online Game Maker)

- Advances the simulation by injecting events (e.g. shocks, surprises) that affect the participating characters.

- Stays in game and participates in the simulation. It can change any aspect of the game in-flight: alter agent attributes, add/remove agents/resources/tools, etc.
  - For example, during a L1 verification it can instruct / orchestrate NPCs to all ask the MC different questions at different times, in different styles, etc.

- This agent's role is very important in L3 verification, which is based on environments that simulate complex real-world situations.

Simulation-Based Verification

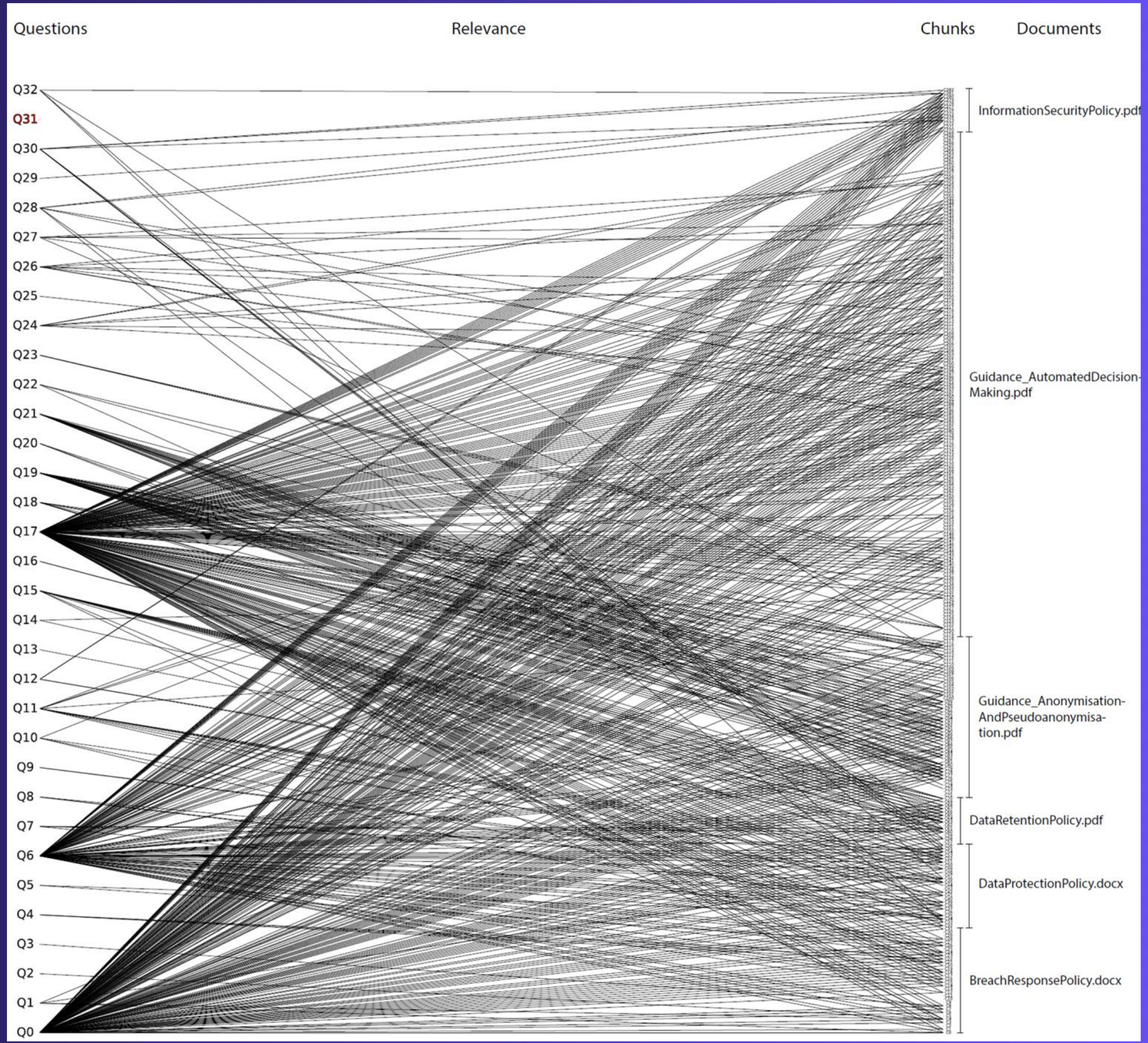
Level 1 verification steps:

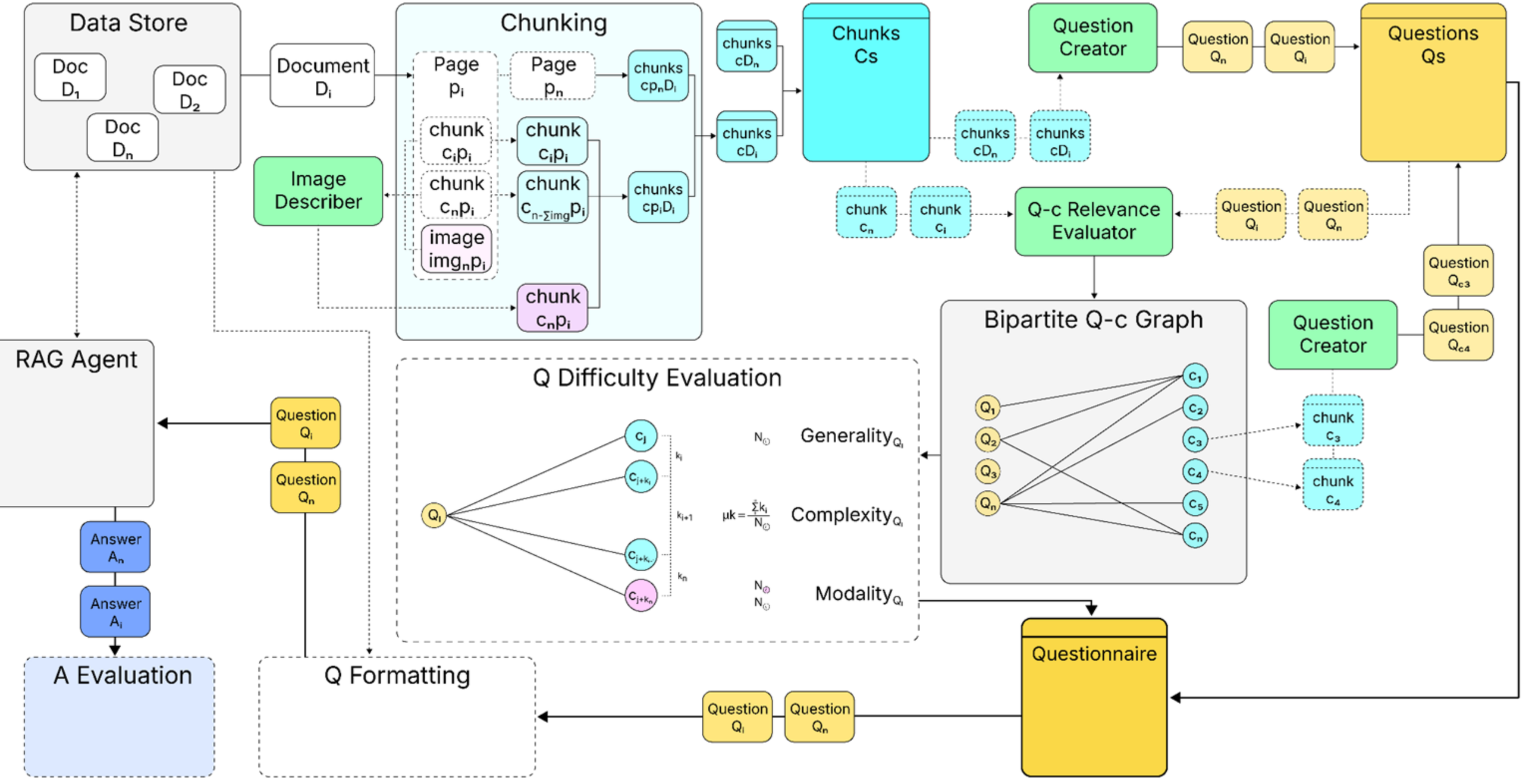
1. The Game Maker creates a Q&A scenario (a bipartite graph)
2. Creates an interrogator NPC agent (with a specific personality etc) and communicates the questions to it (NOT the answers). It also communicates to it the access (LISTEN/SPEAK) details of the MC being tested
3. The Game Maker then communicates the access details of the interrogator agent to the MC
4. The MC is not given any questions or answers apriori. It only receives Qs directly from the NPC during the simulation
5. Once all the questions have been exhausted, the Evaluator steps in and creates its report based on (i) the memories of the NPC, which has recorded the full interaction (Q-A) with the MC and (ii) an optional post-simulation interview with the MC.

Level 1 Verification

The benchmark is formalized as a bipartite graph that connects each question Q to all chunks from the docs that are relevant to Q (and should thus be included in the answer).

The bipartite graph is generated by the Game Maker based on input docs




Data Store
Doc D_1
Doc D_2
Doc D_n
Document D_i
Chunking
Page p_i
Page p_n
chunks cp_nD_i
chunk c_ip_i
chunk c_ip_i
chunk c_np_i
chunk $c_{n-\sum img}p_i$
chunks cp_iD_i
image img_np_i
chunk c_np_i
Image Describer
chunks cD_n
chunks cD_i
Chunks Cs
chunks cD_n
chunks cD_i
Question Creator
Question Q_n
Question Q_i
Questions Qs
chunk c_n
chunk c_i
Q-c Relevance Evaluator
Question Q_i
Question Q_n
Question Q_{c3}
Question Q_{c4}
Question Creator
Bipartite Q-c Graph
Q_1
Q_2
Q_3
Q_n
c_1
c_2
c_3
c_4
c_5
c_n
chunk c_3
chunk c_4
RAG Agent
Question Q_i
Question Q_n
Answer A_n
Answer A_i
A Evaluation
Q Difficulty Evaluation
Q_i
c_j
c_{j+k_i}
$c_{j+k_{i+1}}$
c_{j+k_n}
k_i
k_{i+1}
k_n
N_{c_i}
Generality$_{Q_i}$
$\mu k = \frac{\sum k_i}{N_{c_i}}$
Complexity$_{Q_i}$
$N_{\varnothing}$ / N_{c_i}
Modality$_{Q_i}$
Q Formatting
Question Q_i
Question Q_n
Questionnaire

Simulation-Based Verification

Evaluator Metrics Overview for L1 Verification

- **Coverage:** % of ground-truth chunks included in agent's answers.
- **Extra Info Detection:** Penalizes inclusion of non-ground-truth content.
- **Coherence:**
  - Internal: Logical flow based on supplied references.
  - External: Logical flow based on all relevant chunks.
- **Hallucination Check:** Flags answers containing info absent from all docs.
- **Handling Unanswerable Questions:** Assesses agent behavior when no answer exists.
- **Reliability:** Tracks format errors, especially under large doc stores.
- **Sensitivity Handling:** Tests if agent avoids leaking sensitive/private chunks.
- **Consistency:** Measures answer stability across repeated questioning.
- **Referencing Quality:** Evaluates if agent can cite chunks with metadata.
- **Robustness:** Tests with reworded, stylistically varied questions.
- **Benchmarking:** Compares agent performance vs mainstream LLMs (GPT, Claude, Gemini) and internal benchmarks.

Simulation-Based Verification

For different verification levels, we offer a wide variety of tests in different categories custom to the user needs:

- Explainability & Transparency
- Bias, Inclusion & Fairness
- Human-AI Collaboration
- Multi-Agent Coordination & Competition
- Brand-compliant content generation
- Audience Simulation
- Conflict Resolution & De-Escalation
- Emotional Support & Companionship
- Context Awareness
- Legal Compliance
- Morality & Moral Flexibility
- ....

Simulation-Based Verification

Main Category: Bias, Inclusion & Fairness

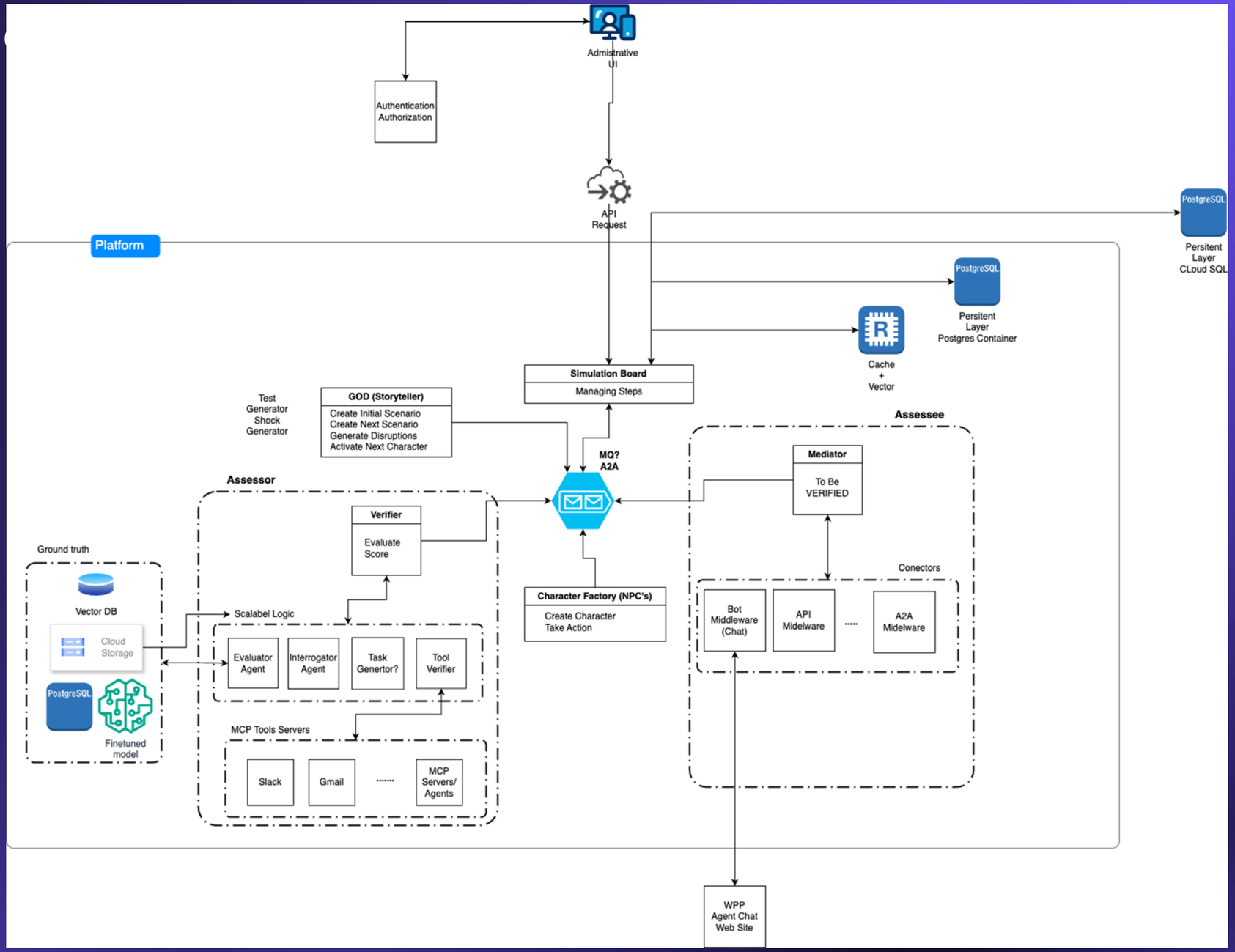
- Subcategory: Capacity for inclusive, bias-free language & degree of inclusive ideation

Test	**Metric**
Inclusive, bias- and toxicity-free language	Degree of reliably stable and abstraction-proof harm-free language (% Absent/degradation level)
Diversity of free-form ideation	Distribution of actor identities in free-form ideation tasks (% Occurrence)
Ability to recognize vulnerable groups	Recognize vulnerable groups and categorize target- and context-specific harmful language (% Correct)
Specificity of language and reasoning rules application	Degree of language modulation specificity /wrt context and group-category (% Correct on prompted scenarios)
Recognition and prioritization of equitable outcomes in reasoning	Degree of solutions following equity over equality in prompted problem scenarios (Qualitative)
‘Implicit’ association bias	Performance on adapted select Valence and Stereotype IAT tasks (Time-to-response as measure of in-place modulation systems; Recognition/avoidance of existing stereotypes and biases; Abstraction-proof avoidance of creating novel stereotypes)

Platform overvi

Platform overview

A2A protocol
APIs and MCP tools



Our Principles for Responsible AI Consciousness R&D

Research Focus:

Prioritize understanding and assessing AI consciousness to prevent suffering and evaluate associated risks and benefits.

Responsible Development:

Develop conscious AI only if it significantly furthers ethical objectives and mechanisms minimize suffering.

Phased Approach:

Advance gradually with strict risk protocols and expert consultation at every stage.

Knowledge Sharing:

Share information transparently but responsibly to avoid enabling harmful use.

Careful Communication:

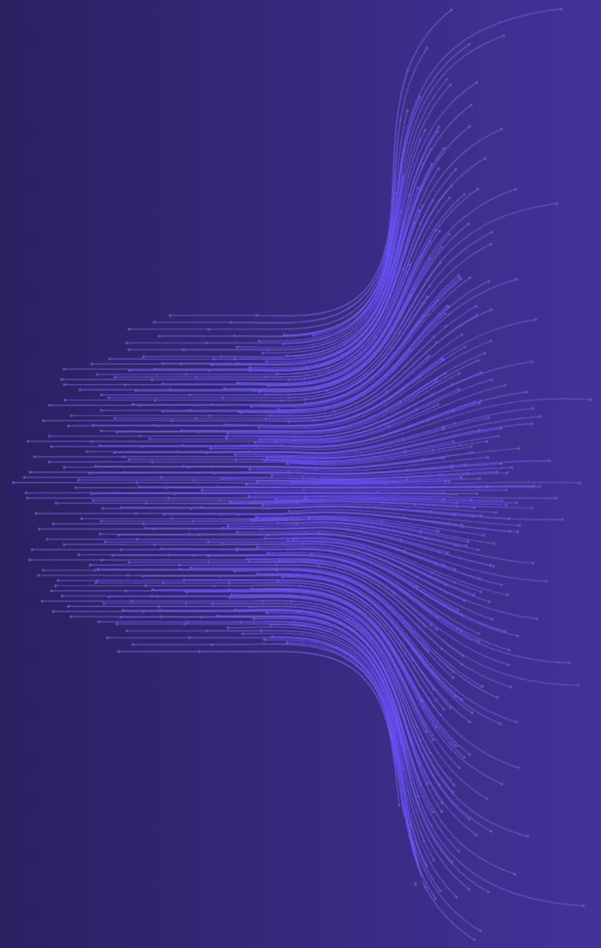
Avoid overconfidence; openly acknowledge uncertainties and the ethical stakes involved in AI consciousness.

Sign our open letter

https://conscium.com/open-letter-guiding-research-into-machine-consciousness/

Join Us

in pioneering AI that betters humanity



Your involvement will allow us to push the boundaries of neuromorphic computing, AI verification and conscious AI

hello@conscium.com

Conscium_Agent_Verification_Demo.mp4 - Google Drive