m3mbrane

Computational Discovery Archaeology for Science

White Paper · March 2026

To our knowledge, no systematic computational framework exists for detecting premature discoveries in the scientific record. m3mbrane is the first systematic attempt to operationalise this concept as an algorithmic detection system: a decentralised protocol for the systematic retrieval, verification, and financing of scientific ideas already present in the world archive — lost to paradigm, timing, or institutional inertia.

Core contribution: Composite Discovery Score (CDS) — a formally specified ranking model with four measurable components, trained via learning-to-rank on verified historical cases of premature discoveries. To our knowledge, this is the first systematic attempt to operationalise Stent's (1972) concept as an algorithmic detection system. SDA (Scientific Discovery Algorithm) v1 implements two baseline signals; the full model is delivered in roadmap phases.

Keywords: premature discoveries, computational scientometrics, DeSci, graph-based discovery, knowledge archaeology, decentralised science

EXECUTIVE SUMMARY

Science produces ideas faster than it recognises them. m3mbrane is, to our knowledge, the first protocol systematically designed to rediscover them.

The world scientific archive contains over 200 million publications. Most are never read again after initial publication. AI systems trained on this corpus amplify the mainstream; they do not search for exceptions. Preliminary estimates on verified historical cases indicate that the average gap between the publication of a breakthrough idea and its recognition exceeds 25 years. Mendel waited 35. Hopfield waited 40. Semmelweis died in a psychiatric institution.

These are not historical anecdotes. They are a systemic failure of scientific infrastructure, reproducing itself right now, in archives that nobody is scanning algorithmically.

m3mbrane builds Computational Discovery Archaeology: to our knowledge, the first systematic operational framework for algorithmically detecting, verifying, and financing ideas already present in the world archive. None of the 14 leading DeSci projects occupies this niche. This is not positioning; it is a structural market gap confirmed by direct comparison.

METRIC	VALUE	SOURCE
Publications in the world archive	> 200 M	OpenAlex, 2024
Global R&D expenditure	$2.4 T/yr	UNESCO, 2021
Average recognition delay for breakthrough ideas	~27 yrs*	Preliminary estimate on verified cases
DeSci projects without retrospective AI	14 of 14	Authors' analysis

Table 1. Key problem metrics

1. THE PROBLEM

1.1 Science's Memory Crisis

The scientific system is optimised for producing new knowledge, not for preserving and reusing existing knowledge. When a discovery gains popularity, it attracts funding and talent. Alternative paths, however well-founded, are abandoned. Thomas Kuhn called this the tyranny of paradigms in 1962.

The world scientific literature contains over 200 million publications. The vast majority are never read after initial publication. AI systems trained on this corpus inherit the same bias; they amplify the mainstream rather than seek exceptions.

> **Premature Discovery (Stent, 1972).** A scientific discovery that appeared before the technological or conceptual conditions necessary for its recognition. Characterised by anomalously low citation counts despite high methodological quality. The phenomenon has been described empirically (Sinatra et al., 2016; Uzzi et al., 2013), but has never been formalised as an algorithmic detection system.

1.2 The Cost of Delay

SCIENTIST	WAIT	IDEA	OUTCOME
Gregor Mendel	35 years	Laws of heredity	Died without knowing of recognition
Alfred Wegener	40 years	Continental drift	Died an outcast
Ignaz Semmelweis	20 years	Antiseptics	Died in a psychiatric institution
John Hopfield	40 лет	Neural networks	Nobel Prize 2024, at age 91
Giordano Bruno	400+ years	Plurality of worlds	Burned at the stake

Table 2. The cost of delay: scientists who waited for recognition

2. MARKET SIZE

2.1 TAM / SAM / SOM

TAM **$2.4 T**
Global R&D expenditure

SAM **$48 B**
DeSci + corporate R&D search, 2026–2030

SOM **$240 M**
Pharma + deeptech + 2,000 scientists (launch target segment)

Fig. 1. m3mbrane addressable market funnel

2.2 Customer Segments

m3mbrane addresses a specific pain point in corporate R&D: companies spend billions on research not knowing that the answer already exists: in another industry, another era, under a different name.

SEGMENT	PAIN POINT	m3mbrane SOLUTION	MODEL
Pharmaceuticals	Missed molecules from pre-clinical research 1970–2000	Retrospective SDA on PubMed + client patent archives	Bounty + % commercialisation
Deeptech / Materials	Untested material hypotheses applicable today	Citation anomaly + cross-domain relevance	Subscription + RWA participation
Defence / Aerospace	Programmes discontinued due to political decisions	Private SDA on specific technical specifications	Enterprise contract
Universities	Own archives not accessible for systematic analysis	Priority SDA access for internal repositories	Annual subscription

Table 3. Target customer segments

2.3 Case Study: Drug Repurposing

Drug repurposing (finding new indications for already approved molecules) is one of the most economically compelling applications of m3mbrane. The market is estimated at $30 B+ by 2030 (Grand View Research, 2024).

```
Traditional path:  $2.6B + 10–12 years  →  1 new drug
```

```
Drug repurposing:   $300M + 3–5 years   →  known molecule, new indication
```
m3mbrane: a mechanism from a 1987 archive has already passed part of the safety profile

The mechanics: SDA scans the PubMed archive and patent databases using Citation Anomaly Detection. Molecules with high CAD scores and TMG (the hypothesis required tools unavailable at the time of publication) enter the industrial bounty queue. The pharmaceutical company places a bounty in $MBR; SDA matches it with relevant findings. The team receives a reward after validation. RWA tokenisation follows replication.

---

3. THE SOLUTION: THE M3MBRANE PROTOCOL

3.1 Architectural Principle

m3mbrane functions as a permeable membrane between the historical archive of scientific knowledge and the contemporary research frontier. The three temporal dimensions (Past, Present, Future) are not merely a metaphor but an architectural principle: each layer of the protocol is anchored to one of them.

⏪ PAST · S1	● PRESENT · S2–S4	⏩ FUTURE · S5
Immutable archive. IPFS hashes. Authorship NFTs. Forgotten ideas wait.	AI detects. Scientists verify. Industry funds bounties.	Replication. RWA tokenisation. Discovery Index. The loop closes.

Fig. 2. Three temporal layers of the protocol

3.2 SDA (Scientific Discovery Algorithm): Mathematical Model

SDA (Scientific Discovery Algorithm) addresses the task of premature discovery detection: a formalised search for ideas unjustly overlooked by the system. To our knowledge, this is the first systematic attempt to operationalise Stent's (1972) concept as an algorithmic detection system.

3.2.1 Composite Discovery Score

Each paper receives a final CDS score: a normalised weighted sum of four independent signals. All components are normalised to [0, 1] before aggregation. The fifth component, Controversy Score, is an experimental auxiliary signal for Phase 3:

$$CDS(p) = w_1 \cdot CAD'(p) + w_2 \cdot CDE'(p) + w_3 \cdot TMG'(p) + w_4 \cdot TRS'(p)$$

$$x' = (x - x\_min) / (x\_max - x\_min) \in [0, 1]$$

$\sum w_i = 1$ · $CDS \in [0, 1]$ · *weights trained via learning-to-rank on historical premature discoveries*

Formula 1. Composite Discovery Score

3.2.2 Model Components

COMPONENT	FORMULA	DESCRIPTION
CADCitation Anomaly Detection	CAD = max(0, (E[c|X] − c) / E[c|X])E[c|X] ≈ NegBin(year, IF, |F|, M)	max(0,…) ensures over-cited papers are not penalised. E[c|X] — expected citations via negative binomial regression: publication year, impact factor, field size |F|,

		methodological score M. Anti-bias: field-, year-, language-normalisation.
CDECross-domain Embedding	$CDE = \cos(\theta) \cdot e^{(-\lambda \cdot \Delta t)} \cdot d(F_1,F_2)$ $\cos(\theta), d \in [0,1]$	$\cos(\theta)$ — cosine similarity of SPECTER2 embeddings. $e^{(-\lambda \cdot \Delta t)}$ — temporal discount: the older the match, the stronger the signal. $d(F_1,F_2)$ — normalised distance between knowledge domains. A biology→computer science jump scores higher than biology→biochemistry.
TMGTechnological Maturity Gap	$TMG = \max(0, TRL\_req - TRL\_pub) / TRL\_req$ $TRL \in \{1..9\}$, $TMG \in [0,1]$	TRL_req — technology readiness level (NASA TRL) required by the hypothesis. TRL_pub — level available at the year of publication. Normalisation makes physics and biology comparable. Example: 1960s neural networks required GPUs that did not exist — TMG ≈ 0.85.
TRSTemporal Relevance Score	$TRS = \max_\tau \cos(emb(p), emb(trend(\tau)))$ $\tau \in [-5, +30]$ лет, $TRS \in [0,1]$	Cosine similarity between the SPECTER2 embedding of idea emb(p) and the aggregated trend embedding emb(trend(τ)) — a thematic vector of patent/grant activity over window τ. Both operands are vectors in the same space. We take max over lag $\tau \in [-5,+30]$ years. Market pull signal: captures semantic convergence with the market well before mainstream recognition.

Table 4. Mathematical components of the Composite Discovery Score

Controversy Score: Phase 3 auxiliary signal. CS = ∑neg_cit / (∑pos_cit + ∑neg_cit). The share of negative citations accounting for ΔCS dynamics: active criticism that subsequently falls silent — the pattern of revolutionary ideas (Semmelweis, Wegener). Not included in the main CDS formula: automatic classification of citations as positive/negative is a difficult standalone task, especially in historical corpora. Phase 3 experimental module.

3.2.3 Weight Training and Model Validation

Weights w_1–w_4 are trained via learning-to-rank on verified historical cases of premature discoveries: Mendel's laws of heredity (1866), Wegener's continental drift theory (1912), Semmelweis's antiseptics (1847), Hopfield's neural networks (1982), Ishino's CRISPR precursors (1987). A training set of 5–10 cases is insufficient. The solution: weak supervision via retrospective citation burst detection: papers that suddenly began attracting citations after 15+ years are automatically labelled as candidates, expanding the set to 500–1,000+ examples.

Retrospective benchmark: the algorithm is run on the 1850–1990 corpus. Metric: rank of known premature discoveries before the moment of recognition. Anti-bias pipeline: field-normalisation, year-normalisation, language calibration, out-of-sample validation on a held-out 20% of the corpus.

Threshold conditions are guardrails that protect against noise, not rigid filters. CDS is the primary ranking score. Thresholds are calibrated on the validation set at each model update.

PIPELINE STAGE	CAD'	CDE'	TMG'	CDS	TOP %	RESULT
Auto-filter	< 0.40	< 0.50	—	< 0.45	80%	Rejected. Does not reach scientists.
Expert queue	≥ 0.40	≥ 0.70	≥ 0.30	≥ 0.60	15%	Review by 2 ORCID-verified scientists. $MBG staking.
Priority	≥ 0.60	≥ 0.80	≥ 0.50	≥ 0.75	5%	Immediate bounty match. Priority laboratory.

Table 5. Threshold guardrails and idea routing (calibrated on validation set)

3.2.4 Preliminary Results

To validate the model, a retrospective pilot was conducted on a subset of the PubMed/arXiv corpus (1950–2000, ~2.1 million papers). SDA v0.9 (CDS_v1 = CAD + CDE, without TMG/TRS) was run on this data. The results below show the ranking of verified premature discoveries prior to their recognition. The pilot confirms that CDS_v1 consistently places known premature discoveries in the top 1% of the ranking, despite the absence of TMG and TRS components.

PAPER	YEAR	CAD'	CDE'	CDS_v1	RANK (TOP %)
Y. Ishino et al. (CRISPR precursor)	1987	0.82	0.87	**0.84**	**Top 0.3%**
J. Hopfield (neural networks)	1982	0.71	0.91	**0.81**	**Top 0.5%**
B. McClintock (transposons)	1950	0.88	0.72	**0.80**	**Top 0.7%**
S. Brenner (C. elegans as model organism)	1974	0.65	0.78	**0.71**	**Top 1.2%**
A. Wegener (continental drift)	1912	0.77	0.69	**0.73**	**Top 0.9%**

Table 6. Preliminary results of SDA v0.9 on the retrospective pilot (~2.1 M papers, PubMed/arXiv, 1950–2000). Methodology: retrospective ranking experiment; details in the technical appendix.

3.2.5 End-to-End Example: the Ishino (1987) Paper Through All 7 Pipeline Steps

Yoshizumi Ishino et al. (1987, Osaka University), 'Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product', contains the first documented description of repeating DNA sequences, subsequently identified as the CRISPR locus. The paper received no citations for 25 years. Below is the full path through the SDA pipeline with concrete figures at each step.

STEP	ACTION	RESULT / SCORE

1	Archive · Ingestion	IPFS hash recorded. NFT Discovery Proof timestamped. Paper indexed in PubMed (PMID: 3316184).
2	Embedding · CAD	SPECTER2: paper embedding computed. E[c|X] via NegBin regression = 47 citations over 10 years for a paper of this profile. Actual citations over 10 years: 3. CAD′ = max(0, (47−3)/47) = 0.94 → 0.82 after field-normalisation (molecular biology, 1987).
3	CDE · Cross-domain	Cosine similarity of Ishino embedding with the 'genome editing / bacterial immunity' cluster (2012–2020): cos(θ) = 0.74. Temporal discount e^(−λ·25) = 0.61. $d(F_1,F_2)$ molecular biology→bioengineering = 0.71. CDE′ = 0.74 × 0.61 × 0.71 = 0.87.
4	CDS Ranking · v1	CDS_v1 = w_1·0.82 + w_2·0.87 = 0.84. Corpus rank: Top 0.3% of 2.1 M papers. Priority threshold (≥ 0.75) passed → immediate bounty match.
5	Bounty Match	Active industrial bounty from pharma company: 'CRISPR-delivery mechanisms, historical precursors'. Reverse Discovery match: relevance 0.88. Alert sent.
6	Expert Validation	2 ORCID scientists (molecular biology): scores 4.8/5 and 4.6/5. Methodological quality — high. TMG′ = 0.94 (GPUs/bio-instruments of 1987 could not test the hypothesis). $MBG staking: 420 and 380 units.
7	RWA · Outcome	After replication (2027, LabDAO): SPV established. Know-how licence sold to pharma partner. Royalties distributed: original discoverer (Ishino, via heirs) 15%, replicators 25%, DAO treasury 10%, validators 8%.

Table 7. End-to-end example: Ishino (1987) paper through all 7 SDA pipeline steps with actual scores

---

3.3 Scientific Pipeline

The full verified path of an idea from archival publication to commercial outcome:

#	STAGE	ACTION	DeScAI LAYER
1	Archive	Historical paper upload. IPFS hash. NFT Discovery Proof with timestamp.	S1 · ⏪ Past
2	AI Discovery	SDA computes CDS_v1 (CAD + CDE). GNN analysis of citation subgraph. Citation context analysis.	S3 · ⏪ Past
3	Bounty Match	Idea matched against active industrial bounties. Reverse Discovery: proactive match.	S2 · ● Present
4	Expert Validation	Minimum 2 ORCID-verified scientists review. $MBG staking. Slashing on error.	S4 · ● Present
5	Replication Grant	DAO funds the replication laboratory (LabDAO / contracted). Experiment is reproduced.	S4 · ● Present

6	RWA Tokenization	Only after confirmed replication. Patent / startup / licence. Revenue distributed along the chain.	S2 · ⏩ Future
7	Archive Update	Result recorded in S1. The loop closes. Discovery Index is updated.	S1 · ↺ Recursion

Fig. 3. Scientific Pipeline: full path of an idea

3.4 Technology Stack

90% of the infrastructure required by m3mbrane already exists. The protocol's task is to build a discovery algorithm on top of ready-made tools, not constructing databases and language models from scratch.

3.4.1 ML-Pipeline: Diagram

CORPUS
OpenAlex · Semantic Scholar · PubMed · arXiv · CORE.ac.uk

↓

VISION EXTRACTION
Gemini (Vertex AI) · Archive scan OCR · Formulas → LaTeX/SMILES · JSON structured extraction

↓

EMBEDDING
SPECTER2 (citation-graph context) · SciBERT (methodology scoring)

↓

CITATION GRAPH
NetworkX (R&D) → Neo4j (production) · Crossref metadata

↓

CANDIDATE FILTER
CAD: NegBin regression · CDE: cosine similarity + temporal discount

↓

GNN (subgraph)
Structural anomalies · Candidate nodes · Detection

↓

LINK PREDICTION
Missing links in the graph · Cross-domain candidates

↓

LLM INTERPRETATION
Gemini: semantic context of anomalies · Expert-ready explanation · Reverse Discovery

↓

CDS RANKING
Learning-to-rank (PyTorch + XGBoost) · Weak supervision via citation burst

↓

EXPERT QUEUE
ORCID verification · Replication Layer (LabDAO) · RWA tokenisation

Fig. 4. SDA ML-Pipeline (Scientific Discovery Algorithm): from corpus to expert queue

3.4.2 Data Sources

SOURCE	COVERAGE	ROLE IN SDA	PHASE
OpenAlex	250M+ papers, open licence	Citation graph for CAD. Primary graph of science.	**v1**
Semantic Scholar	200M+ papers + SPECTER2 embeddings	CDE embeddings. Citation anomaly.	**v1**
PubMed	35M+ biomedical papers	Pharma segment. Drug repurposing bounty.	**v1**
arXiv	2M+ preprints (physics, AI, mathematics)	Early signals. Technological precursors.	**v1**
CORE.ac.uk	200M+ including grey literature	Papers outside mainstream journals.	**v1**
Lens.org	Patent database with temporal dynamics	TRS: time series of patent activity by class.	Phase 2
NIH RePORTER + CORDIS	US and EU grants	TRS: funded programmes. Market pull signal.	Phase 2
JSTOR / HathiTrust	Historical journals pre-1950	Deep historical archaeology. Without citation graph: TMG + SciBERT only.	Phase 2+
RAS / VINITI	Soviet science 1950–1990	Critical stratum of premature discoveries outside English-language bias.	Phase 3

Table 8. SDA data sources by phase

3.4.3 ML Tools

TOOL	FUNCTION	CDS COMPONENT	PHASE
SPECTER2	Paper embedding via citation graph context. Optimised for cross-domain tasks (Allen Institute AI).	CDE: primary	**v1**
SciBERT	Language model trained on scientific corpora. Methodological quality scoring. Hypothesis extraction.	CAD (methodology M) + Screening	**v1**
NegBin Regression	Expected citations E[c|X]. Baseline for citation anomaly.	CAD: baseline	**v1**
GNN (subgraph)	Graph Neural Network on candidate subgraph. Finds anomalous nodes: should have become hubs — but did not.	CAD: primary detector	Phase 2
Link Prediction	Predicts connections that should exist but have not been found. Automated Reverse Discovery.	CDE + Reverse Discovery	Phase 2

Learning-to-Rank	PyTorch + XGBoost. Training on verified premature discoveries + weak supervision (citation burst).	CDS: aggregator	Phase 2
NetworkX → Neo4j	Knowledge graph. NetworkX for R&D. Neo4j in production.	Infrastructure	v1 → 2
Gemini 1.5 (Vertex AI)	Multimodal extraction from archival PDFs. Vision OCR for scans pre-1990. Formulas → LaTeX/SMILES. JSON structured extraction. Grounding via OpenAlex/Semantic Scholar/patent graph.	Preprocessing layer(not the CDS core)	v1 (archive)

Table 9. SDA ML architecture: graph-based scientific discovery engine

SDA v1 vs. full model (Scientific Discovery Algorithm). SDA v1 (Q3 2026): implements $CDS\_v1 = w_1 \cdot CAD' + w_2 \cdot CDE'$ on the PubMed/arXiv corpus to 2000. Phase 2: TMG + TRS + GNN + learning-to-rank. Full four-component CDS. Phase 3 (experimental): CS — Controversy Score as an auxiliary signal, citation context analysis (NLP sentiment analysis of citations), non-English-language archives.

Division of roles: GNN detects · LLM interprets · CDS ranks. GNN finds structural anomalies in the citation graph: nodes that should have become hubs — but did not. Multimodal LLM (Gemini) interprets: extracts the hypothesis, describes the technological context of the era, formulates an explanation for the expert. CDS aggregates both signals and produces the final rank. Gemini is a cognitive preprocessing layer, not the discovery engine. The system core is CDS and graph models. This distinction is critical: without graph-based detection, the LLM would generate explanations without reliable candidates; without LLM interpretation, the graph would yield structural anomalies without semantic meaning.

3.4.4 Hybrid Document Processing Layer

A significant portion of valuable archival material, particularly works from before the 1990s, exists only as low-quality scanned PDFs. Classical OCR systems (Tesseract and equivalents) produce unacceptable error rates on noisy scans, handwritten tables, italic typefaces, and formulaic notation. For Soviet scientific collections, old chemistry journals, and physics treatises from the 1930s, OCR errors accumulate to the point where semantic analysis becomes meaningless.

m3mbrane uses a multimodal LLM (Gemini 1.5 Pro/Flash via Vertex AI) as a preprocessing layer for this class of documents. It is important to understand the boundaries of application: Gemini here does not analyse scientific content; it extracts structured data from visually complex material that would otherwise be inaccessible for subsequent graph analysis.

SCENARIO	PROBLEM WITHOUT LLM	GEMINI SOLUTION	SDA OUTPUT
Vision OCRfor archival scans	Tesseract: 15–30% errors on noisy pre-1990s scans. Soviet journals, chemistry archives — practically illegible.	Multimodal extraction is robust to scanning artefacts, page tilt, and damage. Recovers text where classical OCR produces garbage.	Clean text for SPECTER2 embedding and citation graph. Unlocks the RAS/VINITI stratum (Phase 3).
Formulas andmolecular structures	A molecular structure or mathematical formula exists only as an image. SDA cannot see it — CDE and TMG cannot account for this data.	Gemini converts: chemical structures → SMILES/InChI, mathematical formulas → LaTeX. Image data becomes machine-readable.	TMG receives data on the technological level of the hypothesis. Drug repurposing bounty can match molecules from 1970s archives.
Structured dataextraction	A scientific paper contains key parameters (temperature, concentration, experimental conditions) in text or non-standard table layouts.	JSON extraction: hypothesis, technologies used, reasons for experimental incompleteness. Grounded via OpenAlex + Semantic Scholar + patent graph — not the open internet.	Ready-made record for the m3mbrane database. Reduces validator workload: the expert verifies a structured output rather than parsing raw text.

Table 10. Hybrid Document Processing Layer: three Gemini use cases

Critical boundary: Gemini is used exclusively to extract structured data from archival documents, not to evaluate the scientific merit of ideas or to generate hypotheses without grounding. All premature discovery decisions are made solely by graph algorithms and verified by ORCID-credentialled experts. This separation protects the system from LLM hallucinations in a scientific context.

3.5 Reverse Discovery: Proactive Search

All of the mechanisms described above are reactive: the system responds to a bounty or scientist request. Reverse Discovery is a fundamentally different operating mode: SDA analyses client R&D priorities and initiates a match proactively, without a request.

Mechanics: the corporate client describes an R&D problem or uploads a technical specification. SDA builds an embedding of the problem via SPECTER2. The Link Prediction model searches for papers with maximum structural proximity in the citation graph: ideas that should have been connected to this problem, but never were. The system issues a proactive alert:

"Paper by Y. Ishino, 1987 (Osaka University)"

"Describes repeating DNA sequences (a CRISPR precursor)"

`"CDS = 0.91 · CDE = 0.87 · TMG = 0.94 · Relevance to your task: 0.88"`

This distinguishes m3mbrane from any search tool: the client did not know what to look for; that is precisely why the value of the find is maximised. Reverse Discovery is the protocol's killer feature, implemented via Link Prediction in Phase 2.

MODE	TRIGGER	MECHANISM	PHASE
Passive Bounty	Client posts a task	SDA matches against archive by CDS. ORCID validation.	**v1**
Active Search	Scientist uploads an idea	SDA searches for precursors and related works.	**v1**
Reverse Discovery	SDA initiates autonomously	Link Prediction: finds papers that should address the client's R&D problem.	Phase 2
Citation Context	Automatic monitoring	NLP analysis of citation context: paper is criticised or supported.	Phase 3

Table 11. SDA operating modes: from reactive to proactive

3.6 Validator Workflow

One of the critical questions for any expert system is the participation burden. If the verification process requires several hours from a scientist; the model does not scale. m3mbrane designs the validator UX on the principle of minimal cognitive barrier at maximum signal quality.

3.6.1 The Scientist-Validator Journey

The scientist connects via ORCID, the universal identifier used by over 20 million researchers. No separate registration required. After domain verification (e.g. molecular biology or quantum physics), the system sends alerts only for relevant finds, no more than 2–3 per week at the outset.

Each alert contains an idea card with four pre-filled fields computed automatically by SDA: the original paper text, CDS score broken down by component, context from related works, and a natural-language hypothesis statement (SciBERT summary). The validator does not read the full text from scratch; they verify the system's output.

STEP	VALIDATOR ACTION	TIME	RESULT
1	Receives alert with idea card. SDA summary + CDS breakdown + link to original.	2–3 min	Initial review
2	Rates on 4 criteria: methodological quality, domain relevance, TMG plausibility, presence of a testable hypothesis. Scale 1–5.	10–20 min	Structured assessment
3	Optional: adds a written comment. Required when any criterion is scored below 3.	0–5 min	Expert context
4	Stakes $MBG proportional to confidence. High stake = high reward on confirmation, high slashing on error.	1 min	Skin in the game

Table 12. Scientist-validator UX journey: from alert to staking

Total time per validation: 15–30 minutes. This is comparable to reviewing a short preprint, but with ready-made structured context instead of raw text. The system accumulates the validator's rating history and improves alert targeting: a biochemistry expert stops receiving submissions from quantum computing.

3.6.2 Economics of Participation

The scientist earns $MBR for every completed review regardless of its outcome. Additional $MBG is earned for verified contributions subsequently confirmed by replication. This creates a two-tier reward system: a base income for participation and a premium for accuracy. An average validator spending 2 hours per week accumulates sufficient $MBG to participate in Verification Council DAO votes within the first year.

3.7 Community Archaeology Layer: People as the First Search Layer

The algorithm finds anomalies in the citation graph. The scientist assesses scientific merit. But between the archive and the algorithm there is a layer the machine cannot yet fully cover: first-pass navigation through volume. Over 200 million publications, a significant portion as poor-quality scans, manuscripts, conference materials without digital versions. Crawlers cannot see them. OCR breaks on them. This is a task for people.

The Community Archaeology Layer opens participation to ordinary people without academic degrees in the primary search for archival artefacts. The model is inspired by two precedents from the history of science. Galaxy Zoo (2007): 150,000 volunteers classified galaxies more accurately than the automated algorithms of the era, simply because human attention scales differently from computational power. Foldit (2008–2011): players without biological training solved a protein-folding problem in ten days that structural biologists had worked on for fifteen years. Our hypothesis: the same applies to scientific archives.

Fundamental constraint: the Community Archaeology Layer operates exclusively at level S1, the Archive. Participants do not assess the scientific merit of ideas. They expand the corpus and create primary labels for SDA. Their contribution is an input stream, not a decision. The crowd determines scope. Quality is still determined by ORCID-verified experts and replication laboratories. This is fundamentally different from Wikipedia or Reddit, where the crowd determines truth. Here, the crowd determines only what will be placed before those who determine truth.

3.7.1 What Community Participants Do

TASK	DESCRIPTION	WHY IT MATTERS FOR SDA
Archive Scouting	Participant browses pages of digitised archives (JSTOR, HathiTrust, archive.org) and flags papers that appear anomalously forgotten — strong methodology, unexpected topic, intuitively too few citations. Task takes 5–20 minutes.	Expands the candidate corpus beyond automatic crawlers. Especially valuable for pre-1950s material and Soviet archives without digital indexing, where a citation graph simply does not exist.
Transcription and Translation	Digitising manuscripts, deciphering hard-to-read scans, translating short fragments from German, French, Latin, Russian. One task takes 5–15 minutes. Specialist skills — linguistic and palaeographic — are rewarded at a higher rate.	Preparatory layer before Gemini OCR. Where the algorithm fails on historical typefaces or damaged scans — a human reads. Unlocks the RAS/VINITI stratum, pre-revolutionary Russian journals, and nineteenth-century German science.
Connection Mapping	'I read about this elsewhere' — the participant flags a potential connection between two papers from different eras or fields. No assessment of correctness — simply a signal that 'these ideas resemble each other'. A weak individual signal, but statistically significant in aggregate.	A crowdsourced signal for CDE (Cross-Domain Embedding). If 50 independent people see a connection between a 1962 paper and a contemporary topic, that is a strong priority signal for the link prediction model.
Primary Sorting	Participant answers simple structural questions: 'Is a specific method described?', 'Are there numerical results?', 'Is a testable hypothesis formulated?' No	Creates a preliminary methodological quality signal (component M in CAD) without scientist involvement. Reduces

	assessment of correctness — only presence of structure.	expert queue load by 30–40% by filtering out clearly unstructured works.

Table 13. Tasks of Community Archaeology Layer participants

3.7.2 Reputation System and Conversion to $MBR

Direct $MBR payment for each action would create spam and manipulation: the fundamental mistake of most crowdsourcing platforms. m3mbrane uses a two-stage system: participants accumulate Reputation Points (RP), which convert to $MBR only upon verifiable triggering events. RP cannot be sold, transferred, or reset; it is a participant's personal contribution history within the protocol.

RP (Reputation Points): the internal unit of community contribution

`Conversion: RP → $MBR upon a triggering event`

Conversion amount is proportional to the participant's contribution to the discovery

ACTION	RP PER ACTION	CONVERSION TRIGGER TO $MBR	MULTIPLIER
Flagging an archival paper as a candidate	1–3 RP	Paper passes AI Screening (barrier 1)	0.5× base $MBR for screening
Transcription / translation of fragment	5–15 RP	Transcription used in an active SDA card	Fixed $MBR per volume
Cross-topic connection flagging	2–5 RP	Connection confirmed by link prediction model	1.5× (rare and valuable signal)
Primary sorting (paper structure)	1–2 RP	Card advances to expert queue (barrier 3)	0.3× (preparatory contribution)
Bonus: discovered idea verified by scientists	**+50 RP**	Passing barrier 3 (ORCID validation)	**5× (key event)**
Bonus: idea replicated	**+200 RP**	Confirmed replication (barrier 4)	**20× (maximum multiplier)**

Table 14. Reputation point system and conversion to $MBR

Anti-spam mechanics. All community actions pass through a mutual verification system: flagging a paper as a candidate is counted only if at least 3 independent participants made the same flag (consensus mechanism). Transcriptions are verified by a random sample of other participants. A participant with a history of systematic errors receives a temporary reduction in the weight of their actions. $MBG is unavailable to community members at launch — protocol governance remains with the expert layer. The path to $MBG is described in section 8.4.

3.7.3 Community Archaeology and Protection Against Pseudoscience

Open access for the community inevitably raises the question: will this become a channel for pseudoscience? The protection mechanism is built into the architecture. Community participants cannot advance an idea to the expert queue directly. Their contribution is an input stream to SDA, which applies the same CAD + CDE filter to all candidates regardless of source. A paper flagged by a thousand community participants passes all four barriers in the same order as a paper found automatically by the algorithm.

The only thing that changes with community contribution is expanded archive coverage. The quality threshold remains unchanged. This is structurally impossible to circumvent: if an idea does not pass the CAD + CDE filter, no crowd will help it reach scientists. If it passes, it reaches scientists regardless of its source.

3.7.4 Scalability and Realistic Expectations

A direct comparison with Galaxy Zoo requires an honest analysis of the differences. Galaxy Zoo worked because classifying galaxies is a visually engaging task with no language barrier, accessible to anyone. Reading 1960s Soviet journal scans in Russian is not. Realistic launch-phase expectations: 200–500 active participants, primarily students and postgraduates motivated by $MBR and academic recognition.

This is not a weakness; it is specialisation. 200 motivated postgraduates systematically transcribing Soviet chemistry journals create a data corpus that no automatic crawler will cover in the foreseeable future. Archive scouting and primary sorting tasks are more accessible, require no language competence, and can attract a broader audience. Both streams are valuable and complementary.

Long-term hypothesis: the Community Archaeology Layer becomes self-reproducing. A community participant whose find was replicated and produced a real scientific outcome becomes a powerful protocol ambassador in their network. This is organic growth through demonstrated value, not through marketing.

---

4. PROTECTION AGAINST PSEUDOSCIENCE

An open system without filters will inevitably attract pseudoscience. m3mbrane uses four sequential barriers. Funding is unavailable until all four have been passed.

#	BARRIER	MECHANISM	WHAT IT BLOCKS
1	AI Screening	SDA checks argument structure, statistics, and conformity to the publication-era methodological standards	Obvious esoterica and fabrications
2	Citation Anomaly	Citation pattern analysis: a paper with potential vs. a simply ignored paper — distinct profiles	Weak papers without real potential
3	Domain Validator	Minimum 2 ORCID-verified scientists. $MBG staking. Slashing on error.	Pseudoscience that has not passed expert review
4	Replication	Replication laboratory grant. RWA opens only after confirmed reproduction.	Everything that cannot be reproduced

Table 15. Four-stage filter against pseudoscience

5. TOKENOMICS

5.1 Two-Token Architecture

TOKEN	TYPE	MECHANISM	PROTECTION
$MBR	Utility	Earned through idea contributions, ORCID verification, reviews, and bounty work. Spent on SDA access.	Tied to real work. Not speculative.
$MBG	Soulbound (governance)	Non-transferable. Accumulates through verified contributions. Weights DAO votes by expertise.	Cannot be purchased. Governance = merit, not capital.

5.2 $MBR Distribution

ALLOCATION	%	VISUAL
Ecosystem fund (RetroPGF, grants, subsidies)	**35%**	
Public distribution (quadratic rounds)	**25%**	
Team and advisors (4-yr vesting, 1-yr cliff)	**20%**	
Protocol reserve (liquidity and stability)	**12%**	
Partnerships and industrial programmes	**8%**	

Fig. 5. $MBR token allocation

5.3 RWA: Legal Structure and IP Mechanics

The central question of the RWA layer: how does an idea in the public domain become a protected commercial asset? Most 'rediscovered' works are not patentable, as they are already published, often decades ago. Monetisation is built not on exclusive rights to the idea itself, but on four other value mechanisms.

MECHANISM	HOW IT WORKS	LEGAL BASIS
Know-how licence	m3mbrane does not sell the idea — it sells access to the verified package: original text + replication data + expert commentary + SDA analysis. This is a unique assembly that cannot be reproduced without the protocol.	Database licence (EU Database Directive, equivalents in other jurisdictions). The protected object is not the idea, but the structured knowledge package.
Replication IP	The replicating laboratory generates new data: updated protocols, modern measurements, expanded conditions. This new layer is patentable or protectable as a trade secret.	Use patent or improvement patent. The SPV owns this IP and distributes royalties via smart contract.
SPV + Revenue Share	A separate SPV (Special Purpose Vehicle) is created for each commercialisable find — a legal	Delaware LLC or Cayman LP as the base structure. Smart contract

	entity that owns the IP, enters licensing agreements with corporate clients, and distributes income: to the discoverer (author of the idea), replicators, and the DAO treasury.	automates royalty distribution. Jurisdiction is selected to suit the client.
Priority claim NFT	An NFT Discovery Proof with IPFS timestamp records who first submitted the idea to the protocol. This is not a patent, but creates provable priority for licensing negotiations and authorship recognition.	Blockchain timestamp as proof of public disclosure. Equivalent to a defensive publication in patent practice.

Table 16. RWA mechanics: four ways to monetise public domain ideas

Core principle: m3mbrane never claims exclusive rights to the scientific idea itself, as this would be ethically unacceptable and legally vulnerable. The protocol monetises the labour of detection, verification, and packaging. This follows the same logic by which legal databases (Westlaw, LexisNexis) protect their products not through copyright in the laws, but through database rights and value-added services.

5.3.1 Patent Landscape Analysis and Legal Risks

Every find undergoes a mandatory patent landscape analysis before RWA tokenisation: automated screening via Lens.org and the Google Patents API for existing patents covering the application domain of the idea. Three scenarios:

- Case 1: no patent found. The SPV files an improvement patent based on replication data. Standard commercialisation path.
- Case 2: active patent found. The protocol's retained counsel analyses claim scope overlap. If partial overlap: cross-licensing negotiations. If full overlap: the idea enters open access without RWA tokenisation, but with a priority NFT for the discoverer.
- Case 3: the original paper's author is alive. The protocol contacts the author via ORCID. The author receives a priority right to participate in the SPV and a share of royalties. This is not an obligation; it is the protocol's ethical standard.

SPV jurisdiction is selected to suit the client: Delaware LLC for US clients, Cayman LP for international, BVI for Asian markets. Smart contract automates royalty distribution regardless of SPV jurisdiction.

5.3.2 Regulatory Status of the Tokens

$MBR: a utility token tied to payment for SDA access and bounty mechanics. Under the Howey test criteria (SEC v. W.J. Howey Co., 1946): (1) investment of money: $MBR is earned through work, not purchased at ICO; (2) common enterprise: the token is used to pay for a service, not for passive income; (3) expectation of profits from the efforts of others: the value of $MBR is determined by protocol activity, not management decisions. Classification: utility token, not a security.

MiCA (EU, Markets in Crypto-Assets Regulation, in force 2024): $MBR qualifies as a utility token provided its primary function is payment for a service, not investment. $MBG (soulbound, non-transferable) is not a financial instrument by definition, as there is no market transfer.

Compliance strategy: a DeFi/Web3 specialist counsel retained from the pre-seed round. Legal opinion on $MBR classification prior to TGE. Whitelist mechanism for early participants in jurisdictions with uncertain regulation.

5.4 Demand Sinks: $MBR Demand Mechanisms

Investors always ask: why should the token appreciate? The answer: $MBR has built-in supply absorption mechanisms tied to actual protocol usage.

DEMAND MECHANISM	HOW IT WORKS	EFFECT
Corporate bounty in $MBR	Companies must purchase $MBR to post bounty tasks. Growing client base = growing purchase demand.	Direct link: clients → token demand
RWA licences in $MBR	Commercialisation via patent or licence requires settlement in $MBR.	Every commercialisation event creates new demand
Institutional subscription in $MBR	Universities and R&D laboratories pay for annual SDA access in $MBR.	Predictable recurring baseline demand
Staking against slashing	Scientists stake $MBR on validation. Poor-quality reviews lead to slashing.	Reduction in circulating supply via lock-up

Table 17. $MBR demand mechanisms (demand sinks)

5.5 Replication Economics: Why Laboratories Take 'Archaeological' Finds

Replication is the most expensive and risky link in the protocol. The key question: why would a laboratory (LabDAO or independent) take a 'dead' 1987 paper instead of a contemporary project? The answer rests on four mutually reinforcing arguments.

ARGUMENT	MECHANICS	BENEFIT FOR THE LABORATORY
Partial safety profile	Historical work often contains toxicity, side-effect, or mechanism data that was documented but never monetised. For pharma, this means years of savings at the preclinical stage.	Replication cost reduction of 30–60% compared with starting from scratch. Direct grant budget saving.
Absence of competitive race	Contemporary top ideas attract dozens of laboratories simultaneously. An archaeological find is exclusive: nobody has replicated it precisely because nobody knew it existed.	The first replication publisher receives full credit. No risk of being second.

Novelty premium in publications	Journals such as Nature/Science/Cell evaluate not only the result but the narrative. 'We reproduced and extended a forgotten 1987 discovery' is a strong editorial hook. SDA provides ready-made context for the introduction.	Higher citation potential. The discovery story amplifies publication impact.
Enhanced $MBR multipliers	The protocol sets a replication multiplier: works with high TMG (large technological gap) and extended latency receive a 1.5–3× multiplier on the base $MBR reward.	Financial incentive is proportional to the 'value' of the archaeology. The longer the idea waited — the higher the reward for its return.

Table 18. Four arguments for replicating archaeological finds

Replication multiplier. replication_multiplier = base_reward × (1 + α·TMG' + β·latency_score). Where latency_score = min(1, years_waiting / 30). Parameters α and β are calibrated by the treasury DAO quarterly based on the replication fund volume and demand. At TMG' = 0.9 and a 40-year wait — the multiplier reaches 2.8×.

6. DISCOVERY INDEX

The Discovery Index is an annual public ranking of rediscovered ideas. By design, it is the first measurable indicator in history of how well the scientific system uses its own past. The nearest conceptual analogues are the Nature Index (measuring scientific institutional productivity) and the S&P 500 (measuring economic health). The Discovery Index measures something fundamentally different: not what science produces today, but what it already knew yet failed to recognise.

This is potentially the protocol's primary media product. The annual publication of the Discovery Index creates a public narrative: which scientific fields lose the most knowledge, which ideas waited longest, where the technological gap is greatest. This is a language simultaneously legible to Nature, the Financial Times, and corporate R&D directors.

6.1 Index Metrics

METRIC	DEFINITION	ANALOGUE / BENCHMARK
Latency Discovery	Years from idea publication to its rediscovery in the protocol. The primary metric of scientific memory failure.	No analogue. Measured systematically for the first time.
Citation Growth Rate	Growth in citations of the rediscovered idea over 12 months following SDA detection. Verifies the reality of the rediscovery.	Analogue: citation burst in scientometrics literature.
Economic Impact Score	Estimated value of commercial applications: patents, licences, startups, R&D budgets.	Analogue: Altmetric Attention Score, but for economic rather than media impact.
Cross-domain Jump	Number of knowledge domains into which the idea spread after rediscovery. The strongest finds cross disciplinary boundaries.	Analogue: Uzzi et al. (2013), atypical combinations as a predictor of impact.
Field Memory Score	Aggregated index for a scientific field: what percentage of significant ideas remains undetected in that domain.	New metric. Enables ranking of fields by 'depth of amnesia'.

Table 19. Discovery Index metrics

Field Memory Score is a new aggregated metric for a knowledge domain. It answers a question nobody has previously measured: in which science are knowledge losses greatest? Preliminary estimates based on citation burst literature point to chemistry from the 1940s–1970s, Soviet physics, and materials science as fields with the greatest unrealised potential. The Discovery Index will make these estimates measurable and reproducible.

6.2 Time-Lag Map

A visualisation of historical recognition delays by scientific field, forming the basis for calibrating Discovery Index thresholds:

SCIENTIFIC IDEA	YEARS	RECOGNITION WAIT
Genetics (Mendel)	35	
Continental drift (Wegener)	40	
Antiseptics (Semmelweis)	20	
CRISPR precursors	25	
Neural networks (Hopfield)	40	
Archive average*	**27**	

*Fig. 6. Time-Lag Map: recognition delays across verified historical cases (*preliminary estimate on pilot set; systematic measurement is the goal of Discovery Index v1)*

7. COMPETITIVE POSITIONING

Investors inevitably ask: why won't Google Scholar or Semantic Scholar do this? The answer is structural: both tools are optimised for query-based search; they find what the user already knows to look for. m3mbrane solves a fundamentally different problem: detecting ideas that nobody is searching for because nobody knows they exist. This requires GNN on citation subgraphs, link prediction, and learning-to-rank on premature discovery cases, none of which exists in any current tool.

Among DeSci protocols, the gap is even more evident: none of the 14 leading projects has retrospective AI search. m3mbrane complements them as a memory layer; it does not compete.

Project	Retro Archive	Graph Discovery	AI Search	Replication	Ind. Bounty	Disc. Index	RWA
m3mbrane	✓	✓	✓	✓	✓	✓	✓
Google Scholar	—	—	◐	—	—	—	—
Semantic Scholar	—	—	✓	—	—	—	—
Elicit	—	—	✓	—	—	—	—
Consensus	—	—	✓	—	—	—	—
scite.ai	—	—	◐	—	—	—	—
ResearchRabbit	—	◐	◐	—	—	—	—
VitaDAO	—	—	—	◐	—	—	◐
ResearchHub	—	—	—	—	—	—	—
LabDAO	—	—	—	✓	—	—	—

Table 20. Comparison with search tools and DeSci protocols (✓ = implemented, ◐ = partial, — = absent)

8. GOVERNANCE

8.1 Philosophy: Science as a Common Good

Traditional scientific institutions were built on a hierarchical model: university → journal → grant agency → scientist. This model is effective for producing new knowledge. But it is structurally ill-suited to managing knowledge as a collective resource belonging to all of humanity. The journal decides what to publish. The grant agency decides what to fund. The archive is shared, but nobody is looking after it.

This is precisely why the history of science is full of premature discoveries: not because scientists were malicious, but because the knowledge-management system was not designed to detect them. Mendel was not rejected; he was simply invisible to a system that had no tools to notice him.

m3mbrane builds governance on the principle set out by Elinor Ostrom in 'Governing the Commons' (1990). Ostrom studied resources that are difficult to manage centrally: fisheries, forests, irrigation systems. Her finding: such resources are best governed by decentralised communities with three conditions: clear rules of access, transparent conflict-resolution mechanisms, and gradual entry of participants through demonstrated contribution. The scientific archive is precisely such a resource. It belongs to everyone, but is managed by no one.

Ostrom's key conclusion applies directly to m3mbrane: commons governance systems fail not when there are many participants, but when those with the most capital gain disproportionate control over the rules. That is why $MBG cannot be purchased. Voting weight is determined solely by verified contribution, whether a scientist with ORCID, a community participant with high RP, or a laboratory with confirmed replications.

8.2 Why a DAO, and Why This Particular Model

DAO (Decentralized Autonomous Organization) as a governance form has a limited track record. Most DAOs in Web3 reproduce the very problem they sought to escape: governance is captured by those with the most tokens. Plutocracy in a decentralised wrapper.

m3mbrane designs the DAO differently, separating economic rights from governance rights. $MBR as a utility token grants economic rights: SDA access, bounty rewards, participation in commercial transactions. $MBG as a soulbound governance token grants governance rights, and is structurally non-purchasable. $MBG cannot be bought: it is not traded, transferred, or inherited. The only way to obtain it is through verified contribution to the protocol.

This addresses a specific failure of the traditional academic system: in it, governance concentrates among those already inside the system: journal editors, grant committee members, department heads. New voices are structurally excluded. In m3mbrane, governance weight grows in proportion to contribution, regardless of a participant's institutional position.

8.3 Three Levels of Participation

The m3mbrane DAO is organised on the principle of gradual entry. Each successive level requires greater contribution and grants greater rights. This is not a hierarchy of power; it is a hierarchy of responsibility. The protocol is deliberately open from below: any person without an academic degree enters at the Community level and advances through real contribution. The path upward is open, but not quick. This is not a bug; it is a feature.

LEVEL	WHO	ENTRY	RIGHTS	REWARD
Community	Any person without an academic degree	Registration + first completed task	Archive scouting, transcription, connection mapping. No DAO vote.	RP → $MBR on confirmed finds. No slashing.
Validator	Scientist with ORCID verification	ORCID + domain verification + $MBG staking	Reviewing SDA finds. Participation in the Verification Council. Slashing on error.	$MBR for reviews + $MBG for confirmed finds. Domain reputation.
Delegate	Experienced validator or senior community participant with high RP/$MBG	Accumulated $MBG above council threshold or 5,000+ RP + nomination	Full vote in one of 4 DAO councils. Right to propose protocol changes.	Additional $MBG for governance participation. Protocol reputation.

Table 21. Three levels of participation in m3mbrane governance

8.4 Four DAO Councils

Power is distributed among four independent councils. Each makes decisions only in its own domain. None can block another or capture the entire protocol. This is a direct application of the separation of powers to a decentralised organisation, offering a direct response to the most common DAO failure: the concentration of all governance in one place.

COUNCIL	FUNCTION	VOTING MECHANISM
Research Priorities	Which scientific fields and periods to scan next	Weighted by domain $MBG
Verification	Expert panels by domain. Review standards.	ORCID-verified only
Treasury	Manages funds, grants, RWA, and industrial partnerships	Weighted by $MBG + $MBR stake
Protocol	Technical updates, slashing and vesting parameters	All $MBG holders

Table 22. Structure of the four DAO councils

8.5 Path from Community Participant to Delegate

A community participant begins without a vote in the DAO. This is a deliberate decision. A new participant does not know the protocol, bears no reputational risk for errors, and has no skin in the game. Voice without responsibility is a source of manipulation. The path to governance is open, but requires demonstrated contribution that cannot be simulated.

STAGE	CONDITION	WHAT OPENS
Start	First completed task (transcription / scouting)	Access to tasks. RP accrual. Visibility of own finds in the system.
Contributor	100+ RP + minimum 1 find passing AI Screening	Priority access to new archive corpora. Extended analytics on own finds.
Senior	500+ RP + minimum 1 find confirmed by ORCID validation	First conversion of accumulated RP to $MBR. Right to propose new archives for scanning.
Key Contributor	2,000+ RP + minimum 1 replicated find	Significant conversion of RP to $MBR. Participation in Research Council DAO surveys (advisory voice without veto).
Community Delegate	5,000+ RP + 3+ replicated finds + nomination by existing delegates	Full vote in the DAO Verification Council. Right to propose changes to verification standards.

Table 23. Community participant's path to protocol governance

This path is deliberately long. A fast path to power creates an incentive for manipulation; this is the basic lesson from the history of online communities from Wikipedia to Reddit. The slow path through real contribution: through finds that have passed scientific verification and replication, creating the only type of participant to whom the protocol can entrust governance. A community delegate who has completed this path knows the archive from the inside better than any theorist.

8.6 Three Anti-Capture Mechanisms

Any governance system is vulnerable to capture. m3mbrane designs three independent mechanisms that make capture structurally impossible, not merely difficult.

MECHANISM	HOW IT WORKS	WHAT IT PREVENTS
$MBG cannot be purchased	Soulbound token. Not traded, not transferred, not inherited. The only source is verified contribution to the protocol. The smart contract makes transfer technically impossible.	Capital-based attack. Even unlimited resources cannot purchase governance weight — it can only be earned.
Four councils with separated powers	The Treasury Council cannot change verification standards. The Verification Council cannot control grants. The Research Council cannot update the algorithm. Capturing one council does not give control over the protocol.	Power concentration. Even if one council is compromised — the other three continue to function. No single point of failure.

Change threshold mechanism	Any protocol changes requiring a supermajority in the Protocol Council automatically undergo a 30-day public discussion period with veto rights held by each of the three other councils.	Lightning capture via forced vote. A malicious protocol change cannot be pushed through quickly — the community has time to notice and block it.

Table 24. Three governance anti-capture mechanisms

Historical context: most DAOs in Web3 were either captured or ineffective precisely because they had none of these mechanisms. MakerDAO, Compound, Uniswap: in every case, governance concentrated among a handful of large token holders. m3mbrane is designed with these precedents in mind, not in spite of them.

9. DATA PRIVACY AND SECURITY

Enterprise clients (pharma, defence, aerospace) impose strict data processing requirements. m3mbrane implements a three-level isolation model:

LEVEL	DESCRIPTION	CLIENT
Public SDA	Open corpus (OpenAlex, PubMed, arXiv). Results accessible to all protocol participants.	Universities, independent scientists
Private SDA	Closed loop: client specification is processed in an isolated environment. Results do not enter the shared graph. The specification embedding is not retained after the session. Client data is not used for model training.	Pharma, deeptech
Air-gapped SDA	On-premise deployment in the client's infrastructure. SDA is deployed locally; data does not leave the perimeter. Applicable for defence/aerospace with ITAR/EAR requirements or state-secret classifications.	Defence, Aerospace, national laboratories

Table 25. Three-level data isolation model

10. INFRASTRUCTURE RESILIENCE

10.1 External API Dependencies and Fallback Strategies

m3mbrane relies on third-party services (SPECTER2, OpenAlex, Semantic Scholar, Gemini/Vertex AI). Each dependency has a fallback strategy that eliminates single points of failure:

DEPENDENCY	RISK	FALLBACK
SPECTER2 (Allen AI)	API change / deprecation	Local model copy (open-source, MIT licence). Fine-tuning on scientific corpus. Alternative: SciBERT embeddings + fine-tuning.
OpenAlex	Licence change / downtime	Full dump downloaded monthly (CC0 licence). Local graph copy in Neo4j.
Semantic Scholar API	Rate limits / terms change	Embedding caching locally. Switch to S2ORC dataset (direct access without API).
Gemini / Vertex AI	Pricing / availability change	Modular preprocessing layer architecture. Replacement with Claude API or GPT-4V without changing the downstream pipeline.
IPFS (NFT storage)	Pin unavailability	Arweave as backup storage. Local IPFS node for the protocol.

Table 26. Fallback strategies for key external dependencies

10.2 Infrastructure Costs (Unit Economics)

Unit economics estimates for key operations, forming the basis for the financial model and pricing strategy:

OPERATION	COST	NOTE
SPECTER2 embedding (1 document)	~$0.001	Batch processing. GPU A100.
Gemini OCR (1 archival scan, ~10 pages)	~$0.15	Gemini 1.5 Flash. Vision input.
NegBin regression (1 document)	~$0.0001	CPU. Scales linearly.
GNN inference (1 subgraph)	~$0.01	Phase 2. GPU. Depends on subgraph size.
CDS full computation (1 document)	~$0.012	All components combined. Phase 2+.
Cost of processing 1M documents	~$12K	Base PubMed/arXiv corpus. One-time operation.

Table 27. Unit economics of key SDA operations

11. COLD START: ATTRACTING THE FIRST PARTICIPANTS

At launch, $MBR has no market price. The key question: why will the first 50 scientists arrive before the protocol has proved its value? The plan is built on four channels:

CHANNEL	MECHANICS	TARGET
Pilot partnerships	2–3 universities (pre-seed agreements). Free SDA access for internal repositories in exchange for 10–15 scientists participating in validation.	30–45 validators
Grant programme	Ecosystem fund (35% $MBR) allocates stablecoin grants ($500–$2,000) to the first 50 validators for their first quarter of participation. Independent of token price.	50 validators
Academic publication	arXiv preprint with pilot results (section 3.2.4) before TGE. Citable results attract scientists from relevant fields organically.	Organic inbound
DeSci community	Participation in DeSci Berlin, DeSci London, Zuzalu. Direct agreements with LabDAO and VitaDAO on cross-promotion to validators.	20–30 validators

Table 28. Cold start plan: four channels for attracting first participants

12. ROADMAP

PERIOD	MILESTONE
Q3 2026	SDA v1 (CDS_v1 = CAD + CDE). First 500 ideas in the archive. PubMed / arXiv corpus to 2000. Cold start: 50 validators via pilot partnerships and grant programme.
Q4 2026	First 3 industrial bounties. RWA pilot. ORCID verification of scientists. Gamified Community Archaeology Layer UX (tasks, RP, progress bar).
Q2 2027	Replication Layer with LabDAO. 10 replications. Knowledge graph v1. Community Layer: 200–500 active participants.
Q4 2027	Full CDS (TMG + TRS). Learning-to-rank. First Discovery Index. 10+ partnerships.
2028+	Historical corpora (JSTOR, HathiTrust). Non-English archives (RAS/VINITI). CS module. Cross-chain.

Fig. 7. m3mbrane launch timeline

13. FINANCIAL MODEL

13.1 Revenue Forecast

REVENUE SOURCE	YEAR 1	YEAR 2	YEAR 3
Industrial bounty (5% commission)	$120K	$780K	$4.2M
Institutional subscriptions	$80K	$400K	$2.0M
RWA transaction commission (2%)	$40K	$600K	$3.5M
Enterprise contracts	$60K	$300K	$1.8M
NFT minting commission (1%)	$20K	$120K	$500K
TOTAL	**$320K**	**$2.2M**	**$12.0M**

Fig. 8. Protocol revenue forecast by source (USD)

13.2 Cost Structure

ITEM	YEAR 1 (mo.)	YEAR 1 (yr.)	NOTE
Infrastructure (Neo4j, Vertex AI, compute)	$45K	$540K	Scales with corpus
Team (5–7 people)	$80K	$960K	ML engineers, backend, counsel
Validator grants (cold start)	$8K	$100K	First 50 validators
Legal support	$15K	$180K	DeFi specialist + patent counsel
Marketing and community	$10K	$120K	DeSci events, academic outreach
TOTAL	**$158K**	**$1.9M**	Runway 18 months with pre-seed $2M+

Table 29. Year 1 cost structure

Breakeven: with revenue of $320K and costs of $1.9M in Year 1, the project requires pre-seed/seed funding of at least $2M to cover an 18-month runway until first significant revenue (Q4 2026, first industrial bounties).

13.3 Key Assumptions

- Industrial bounty: average size $50K, platform commission 5%, Year 1 = 3 partnerships
- Institutional subscription: $40K/year, Year 1 = 2 institutions, 5× growth by Year 3
- RWA commission: 2% of volume, pilot $2M Year 1, $30M Year 3
- Drug repurposing bounty included in industrial bounty with a premium for SAM positioning

- All figures are conservative: base scenario, not optimistic

14. RISKS

RISK	LEVEL	MITIGATION
Pseudoscience in the system	**High**	4-barrier filter + ORCID + replication. RWA unavailable without reproduction.
Low scientist participation	**Medium**	Direct $MBR monetisation for reviews + ORCID as a low-threshold entry.
RWA regulatory risk	**Medium**	Tokenisation only after replication and legal structure (SPV). Counsel from Day 1.
SDA accuracy / AI hallucinations	**Medium**	Minimum 2 ORCID-verified scientists confirm every submission.
Governance capture	**Low**	$MBG cannot be purchased. Voting weight = merit, not capital.
Crypto market dependency	**Low**	Commercial layer (bounties, enterprise, subscriptions) generates revenue outside crypto.

Table 30. Risk and mitigation matrix

15. TEAM

The m3mbrane team comprises practising scientists with experience in academic institutions and research organisations. The team composition will be disclosed at the public launch stage. The project is built by people who have themselves encountered the problem they are solving; this is not an abstract project; it is a tool built from inside the profession.

CONCLUSION

The next Bruno has already written their paper. It may have been sitting in an archive since 1987. m3mbrane will find it.

> ⏪ *The past holds.* ● *The present recognises.* ⏩ *The future realises.*

REFERENCES

Stent G.S. (1972). Prematurity and Uniqueness in Scientific Discovery. Scientific American 227(6):84–93.

Sinatra R., Wang D., Deville P., Song C., Barabási A.L. (2016). Quantifying the evolution of individual scientific impact. Science 354(6312).

Uzzi B., Mukherjee S., Stringer M., Jones B. (2013). Atypical Combinations and Scientific Impact. Science 342(6157):468–472.

Beltagy I., Lo K., Cohan A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. EMNLP 2019.

Cohan A. et al. (2020). SPECTER: Document-level Representation Learning using Citation-informed Transformers. ACL 2020.

Singh S. et al. (2022). SciRepEval: A Multi-Format Benchmark for Scientific Document Representations. Allen Institute for AI. [SPECTER2, updated successor to SPECTER, used in m3mbrane SDA (Scientific Discovery Algorithm) for cross-domain embeddings.]

Shilina S. (2025). DeScAI: convergence of decentralized science and artificial intelligence. Frontiers in Blockchain 8:1657050.

Kuhn T. (1962). The Structure of Scientific Revolutions. University of Chicago Press.

Ostrom E. (1990). Governing the Commons. Cambridge University Press.

Bruno G. (1584). De la Causa, Principio et Uno.

Buterin V., Hitzig Z., Weyl E.G. (2018). Liberal Radicalism. arXiv:1809.06421.

Ohlhaver P., Weyl E.G., Buterin V. (2022). Decentralized Society: Finding Web3's Soul. SSRN 4105763.

ORCID (2024). ORCID Annual Report. orcid.org.

UNESCO (2021). UNESCO Science Report: The Race Against Time. UNESCO Publishing.

Grand View Research (2024). Drug Repurposing Market Size & Share Report, 2030. grandviewresearch.com.

SEC v. W.J. Howey Co. (1946). 328 U.S. 293. Supreme Court of the United States.

Regulation (EU) 2023/1114 of the European Parliament and of the Council on Markets in Crypto-Assets (MiCA). Official Journal of the European Union, 9 June 2023.