# Cairn - whitepaper

## Problem formulation

The current structure of scientific research suffers from several systemic issues that limit long-term impact and accessibility:

- **Poor Incentives for Impactful Work**: Academic recognition and funding systems tend to reward novelty and publication volume over real-world usability, long-term impact, and reproducibility.
- **Reproducibility Crisis**: A large portion of published research cannot be reliably reproduced, leading to wasted effort, erosion of trust, and slower scientific progress.
- **Centralized Funding Gateways**: Funding decisions are concentrated in a few institutions, which can limit diversity in research agendas and hinder innovation from underrepresented contributors.

As AI advances rapidly, **embodied intelligence**—the ability of AI systems to perceive, reason, and act in the physical world—has emerged as a crucial frontier. However, the structural issues in science are even more acute in this field due to its high commercial potential and technical complexity.

The increased pressure for rapid, commercially relevant results in embodied AI research often leads to **reduced transparency**, limited openness, and shortcuts in reproducibility. This undermines scientific rigor and slows collaborative progress across the field. At the same time, the overwhelming volume of research papers—many of which have not been independently reproduced—makes it difficult for funders and stakeholders to **identify truly impactful** contributions.

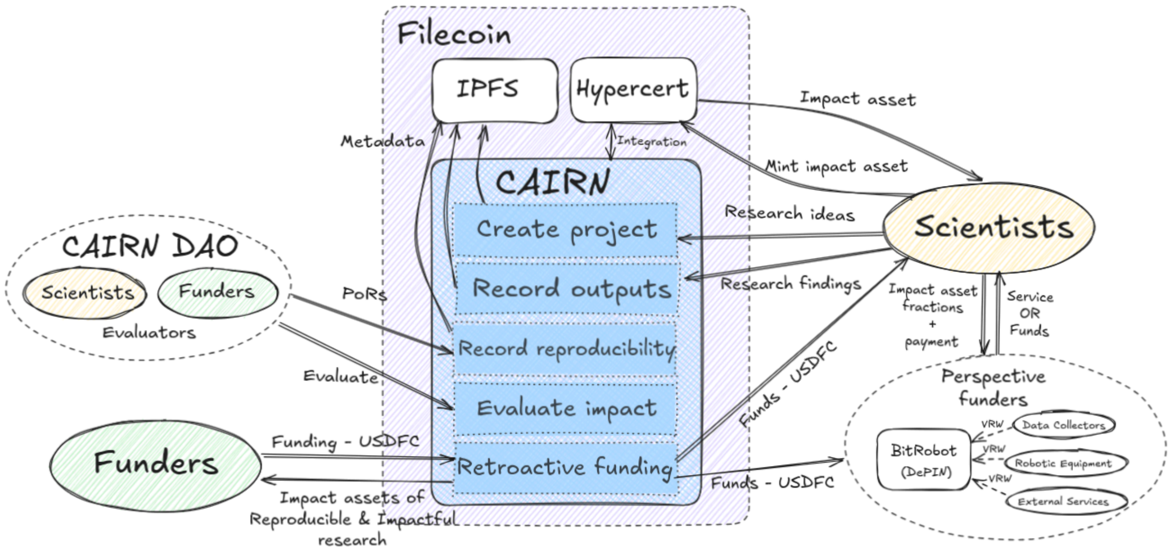
A major issue is the lack of visibility into how scientific outputs such as datasets, simulators, algorithms, or hardware setups are reused and built upon. Without a systematic way to trace this reuse, it becomes nearly impossible to assess long-term influence or reward foundational work. Compounding this, the reproducibility of embodied AI research is inherently more complex than in other AI domains. Reproducing results often requires replicating physical setups, managing multimodal sensor data, implementing real-time control systems, or navigating intricate simulation environments. These technical and logistical challenges make reproducibility expensive and often impractical.

Together, these factors create a research ecosystem where reusable and impactful work is difficult to track, undervalued in funding decisions, and ultimately underutilized—slowing progress in one of the most promising and societally significant areas of artificial intelligence.

# CAIRN

> 💡 A protocol that tracks the reuse and reproduction of scientific outputs in the field of embodied AI, and retroactively reward impactful work (both scientific and supporting). Transparency that leads to reproducibility is rewarded if it is also impactful.

The system is designed to incentivize and validate impactful, reproducible scientific research by connecting scientists, funders, and decentralized infrastructure via a Web3-native workflow. It is built around the **Cairn platform**, with various stakeholders and components described below:




## Scientists

Scientists contribute research outputs as the primary producers in the system. These outputs can include:

- Machine learning models and architectures,
- Annotated or raw datasets,
- Benchmark setups,
- Proofs of Reproducibility (PoRs).

## Funders

Funders provide capital to support the research process. They may contribute:

- Upfront (prospective) funding for new projects,

- Retroactive funding after projects have demonstrated impact and reproducibility.

Funders rely on the Cairn DAO for impact evaluation.

## External Services

External services support the scientific process by providing:

- **Data collectors** – crowd-sourced or automated data acquisition agents,
- **Robotic equipment** – physical infrastructure for experimentation or deployment,
- **Teleoperators / Human participants** – assist with remote operation or data validation.

**BitRobot** is an example of DePIN which provides Virtual Robot Workers (VRWs) on-demand in exchange for payment.

## Hypercert

A **Hypercert** is a tokenized certificate of impact that represents the value and significance of a scientific contribution. The Cairn platform mints Hypercerts as **impact assets**, which are then:

- Credited to contributing scientists,
- Used by funders to track the return (in terms of societal or scientific value) on their investments,
- Traded or used as proof of contribution in broader reputation or funding ecosystems.

## Filecoin

The Cairn platform integrates Filecoin technology as a foundational layer to ensure decentralized, secure, and verifiable storage of scientific research data and metadata. Smart contracts are deployed on the **Filecoin network** to manage transparent and immutable interactions between stakeholders. All research metadata—including machine learning models, datasets, benchmarks, and Proofs of Reproducibility (PoRs)—is stored on **IPFS** through **storage onramps** that interface seamlessly with Filecoin's decentralized storage network. This guarantees data provenance, integrity, and long-term availability. Additionally, the **USDFC stablecoin**, native to the Filecoin ecosystem, is used as the platform's funding currency, enabling stable, trustless, and efficient capital flows between funders and scientists.

## Cairn Platform

The **Cairn platform** is a Web3-enabled application that orchestrates the research lifecycle and value attribution. Its core modules allow users to:

1. **Create projects** – initiating new research efforts.

2. **Record outputs** – publishing models, data, and results.
3. **Record reproducibility** – documenting verification steps.
4. **Evaluate impact** – measuring downstream value and reproducibility.
5. **Distribute retroactive funding** – rewarding impactful work.

### Cairn DAO

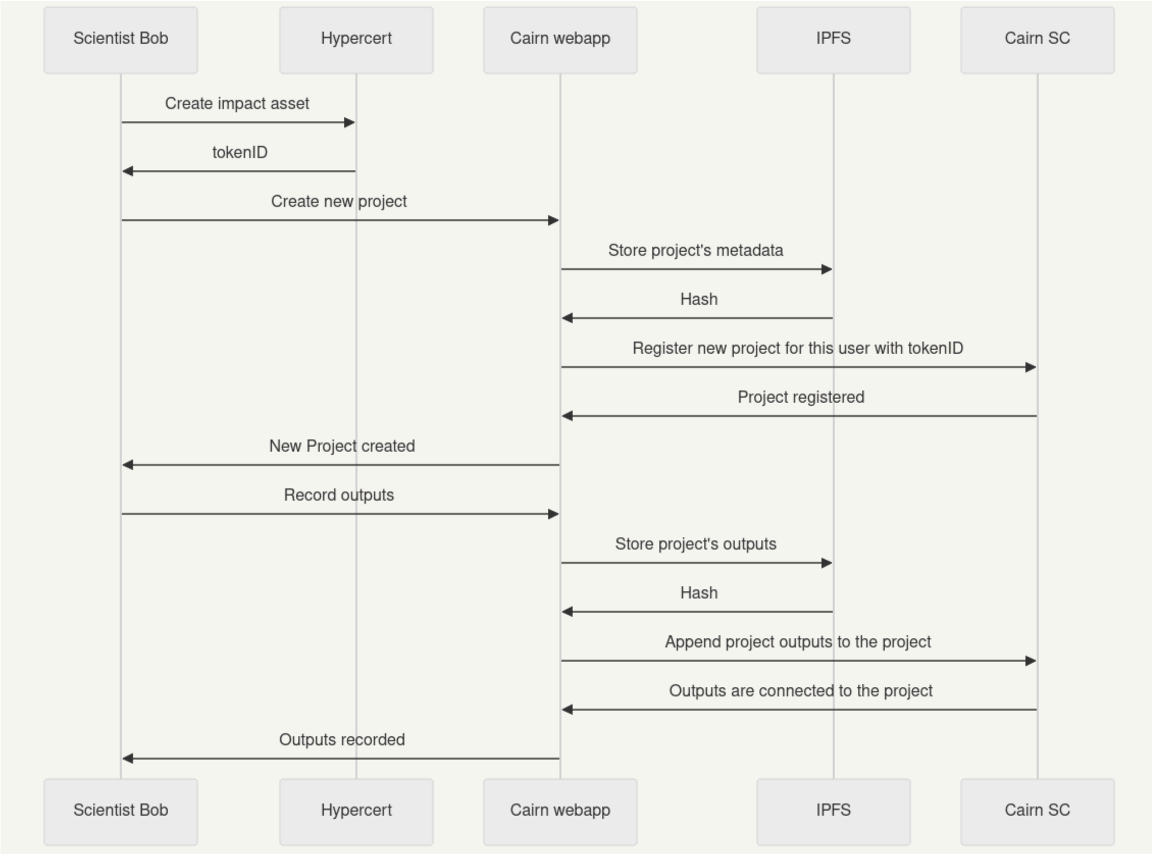
The **Cairn DAO** is a decentralized governance body composed of scientists, funders, and evaluators. It plays a critical role in:

- Voting on evaluations of research impact,
- Deciding on funding allocations,
- Governing platform evolution and integrity.

## Processes

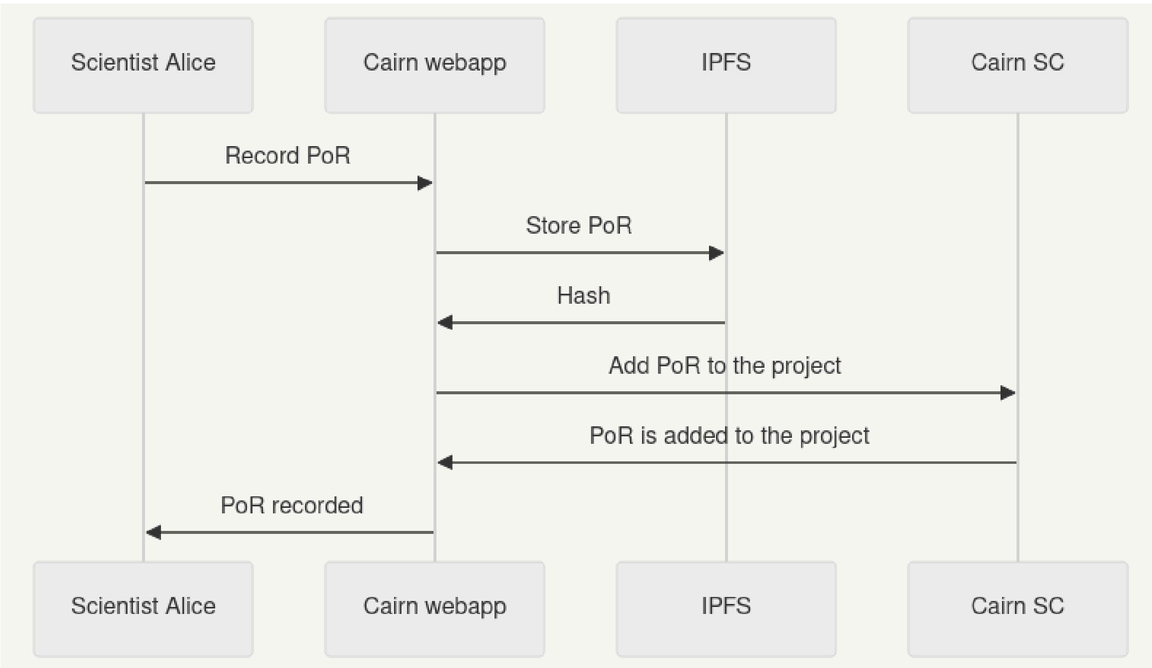
- Create project + Record outputs

  This flow shows how a scientist registers a new research project on the Cairn platform. Scientist must register with decentralized identity (DID) and provide a Proof of Humanity (PoH) for authentication. Scientists who have contributed a certain number of **Proofs of Reproducibility (PoRs)** become eligible to create new projects. The project metadata and outputs are stored on IPFS, registered on-chain via smart contracts. Before project is registered on smart contract impact assets must be minted using Hypercerts. These assets represent the scientific contribution and make the project eligible for future funding.



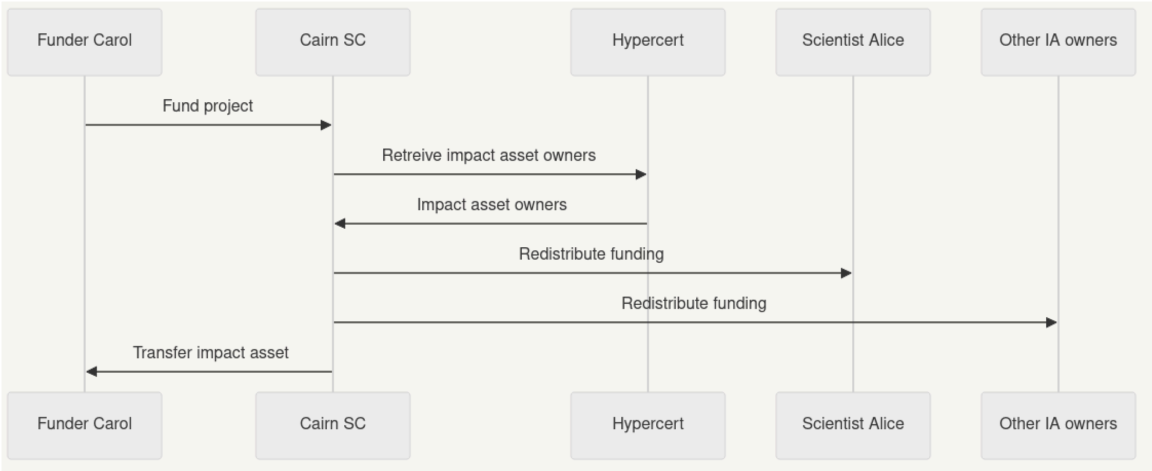

- Record reproducibility

  This process illustrates how scientists or verifiers submit a PoR for an existing project. The PoR is stored on IPFS and linked to the project via smart contracts, improving the project's credibility and eligibility for retroactive funding.




- Evaluate impact + retroactive funding

This flow depicts how a funder allocates retroactive funding to a completed project. The smart contract retrieves the owners of the associated Hypercert impact assets and redistributes funds proportionally, rewarding contributors based on verified impact. The impact asset fractions are all transferred to the funder of the project.




# Proof-of-reproducibility

The **Proof of Reproducibility (PoR)** system is designed to provide verifiable evidence that a scientific experiment has been successfully reproduced by a third party. It is an integral part of the platform, ensuring that the replication process is transparent, trustworthy, and resistant to tampering. This technical framework combines **cryptographic hashing**, **decentralized storage (IPFS)**, and **smart contracts** to manage and validate reproducibility evidence. The number of reproduction is limited as confirmed and well known methods do not need constant reproduction.

## 1. Data Structure for PoR

The PoR is composed of multiple components that capture different aspects of the reproduction process. These components are stored securely and immutably using **IPFS (InterPlanetary File System)**, and each piece of proof is hashed and linked to a **smart contract** to ensure accountability and non-repudiation.

The following data is included in the PoR:

- **Link to Experiment Outputs**:
  - The outputs of the replicated experiment (e.g., robot behavior, simulation results) are captured and stored in **IPFS**. These outputs represent the **final results** of the replication process, and they provide the core evidence that the experiment has been performed correctly.
  - The **IPFS link** to these outputs is generated and stored along with other metadata for traceability.
- **Reproduced Results**:

    - The results of the replication (e.g., accuracy, performance metrics, final state of the experiment) are stored in **IPFS** as data files or logs. These results should match the expected outputs of the original experiment.
    - **Hashing** of these results and storing them securely ensures that the replicator's performance is verifiable and immutable. The **hash** of this data is stored in the associated **smart contract** for easy validation.
- **Logs (Code Outputs + Timestamps + Hashes, Simulator Logs)**:
    - The replicator provides detailed **logs** of the entire experiment execution, including:
        - **Code outputs** (e.g., simulation results, robot movements, or task results).
        - **Timestamps** for each experiment step to show the timeline of the experiment.
        - **Hashing of log files** to ensure their integrity and prevent tampering.
        - **Simulator logs** (if applicable), which capture system states, sensor feedback, and other metadata crucial for verifying that the replication process followed the described setup.
    - These logs are **hashed** and stored on **IPFS**. The hash of the log data is then registered on the **smart contract**, linking the experimental proof to the contract.
- **Receipts from External Service Providers**:
    - If the replication process involves **third-party service providers** (e.g., **DePIN** providing sensor data, robotic equipment, or cloud compute services), receipts or **proofs of service utilization** are generated.
    - These receipts are mapped to **service provider logs** (e.g., usage logs, data timestamps) and stored on **IPFS**.
    - The **hash** of the receipts is stored on the **smart contract** to demonstrate that the replicator correctly utilized the external services in accordance with the original experimental setup.
- **Hardware-based Proof (Video + Watermark + Hash + Timestamp)**:
    - For **hardware-based experiments**, **video footage** of the experimental process is provided as evidence that the replication was physically performed.
    - The video is **watermarked** with a unique **Proof ID** to ensure that it is uniquely linked to the experiment and cannot be reused for other replications.
    - A **cryptographic hash** of the video file is generated, and the hash is stored on **IPFS** for immutability. The video's **timestamp** is also recorded to prevent tampering.

- The **video proof** includes a visible **watermark** that contains the **PoR ID** and **metadata** such as experiment type, equipment used, and replicator identity. This ensures that each video can be independently verified as authentic and relevant to the specific proof.

---

## 2. Storage and Linking on IPFS

All the data related to the **PoR** (outputs, logs, receipts, video footage) is stored on **IPFS** to ensure **decentralized, permanent storage**. IPFS provides the following benefits:

- **Decentralization**: Data is stored across multiple nodes, ensuring that no single entity controls or can alter the data.
- **Immutability**: Once data is uploaded to IPFS, it is assigned a unique **content-addressable hash** that guarantees data integrity. This ensures that the evidence cannot be altered after submission.
- **Efficiency**: IPFS allows for efficient retrieval of large files such as video footage and simulation logs, ensuring that data can be accessed quickly during the verification process.

Each file stored on IPFS will be indexed with a unique **CID (Content Identifier)**, which acts as a pointer to the file's content. These CIDs are used to link the data to the corresponding **smart contract**.

---

## 3. Ensuring Integrity and Preventing Fraud

To prevent fraud and ensure the validity of the PoR:

- **Cryptographic Proofs**: Each PoR element is **hashed** (video, logs, results) before being stored on IPFS. The cryptographic hash guarantees that the data cannot be modified without detection.
- **Watermarking**: The video evidence is watermarked with a unique **Proof ID** to prevent the reuse of the same video footage in multiple experiments.
- **Smart Contract**: The **smart contract** serves as an immutable record of all registered PoRs. It ensures that any change in the experiment or results would invalidate the previous proof and prompt a new validation process.

---

## 4. Dispute Resolution and Human Review

While the system is designed to be **optimistic**, disputes may still arise regarding the validity of a PoR. The following process handles disputes:

1. **Dispute Filing**: If a participant believes that the proof is invalid (e.g., evidence has been tampered with, experiment conditions were not followed), they can file a dispute.

2. **Improved PoR Submission**: The disputing party can submit additional evidence to support their claim.
3. **Human Intervention or Voting**: If the dispute remains unresolved, the platform facilitates **human intervention** (e.g., a live demonstration via web call) or allows **voting** by trusted community members to resolve the dispute.
4. **Final Decision**: The outcome of the dispute is recorded on the smart contract, ensuring that all parties are aware of the final verdict.

# Retroactive funding

The platform introduces a **Retroactive Funding Mechanism** to incentivize high-quality, reproducible scientific work in the field of **embodied AI**. Rather than issuing grants up front, funding is distributed **after a project demonstrates measurable outcomes** — particularly:

1. **Tangible outputs**,
2. **Proven reproducibility** (via PoRs from independent scientists), and
3. **Impact assessment** as determined by community governance.

This model aligns scientific incentives with rigor, transparency, and real-world contributions — rewarding **what works**, not just what is proposed.

## Components of the Retroactive Funding System

### 1. Eligibility Criteria

A scientific project becomes **eligible for retroactive funding** when it meets all of the following:

- ✅ **Outputs Submitted**

  The project has submitted artifacts such as datasets, simulators, code, or hardware results — stored immutably (e.g. via IPFS) and referenced on-chain.

- ✅ **Proofs of Reproducibility (PoR)**

  The project has received a threshold number of valid PoRs submitted by third-party scientists.

- ✅ **Impact Evaluation Completed**

  The project must undergo a **community-based impact evaluation**, which determines the magnitude of its contribution and its eligibility tier for funding.

---

### 2. Impact Evaluation Mechanism

Evaluation is performed by a **DAO** composed of scientists, funders and other Cairn community members. The DAO votes on each eligible project's **impact category**, which determines the size and timing of retroactive funding.

**Possible impact tiers**:

| Tier | Description | Funding Share |
|---|---|---|
| 🔷 Low | Incremental or niche contribution | Small |
| 🔶 Medium | Significant improvement or application | Moderate |
| ♦️ High | Groundbreaking insight, large-scale utility | High |

## 3. Funding Flow

Retroactive funds can be sourced from:

- DAOs, VCs, or ecosystem funds
- A **treasury funded by public goods grants**

Once a project is evaluated:

1. A **funding amount is calculated** based on the impact tier.
2. The funds are distributed to all the owners of impact assets (authors of the project, prespective funders and external service providers - DePIN that are payed with impact assets)