Sustainable Infrastructure for AI-Empowered Open & Collaborative Decentralized Science

Dominikus Brian 钟鸿盛 | domi@dreambrook.tech
www.dreambrook.tech

Decentralized science (DeSci) movement have been going around for quite long. Beginning from the advent of Wikipedia in which knowledge and peer-to-peer knowledge curation become widespread, there has been a great drive to propel decentralized science to next level. ResearchHub is another platform that has promote sharing, peer-reviewing, and summarizing research documents. Empowered with blockchain technology and other digital infrastructure that is maturing over the decades, now more should and can be realized for Decentralized Science.

As in any endeavor in life invention or initiative are usually composed of two main components: Idea and Action. The idea part of Decentralized Science is been growing strongly and there are great many ideations for amazing initiative. However, in the action side, the movement often not moving forward due to either technical barrier, lack of incentive, or simply insufficient resources to support it. All this come to one word, Infrastructure. Meaning there's still lack of ready to use infrastructure to enable these idea flourish.

Entering through 2020s and beyond Artificial Intelligence (AI), initially dominated by mere machine learning (ML) that predict or forecast particular output based on input data being learned, now have move forward toward the creation of various intelligence agent, spanning from Large Language Models, to autonomous AI-Agents, Smart IoT, and Advanced AI-powered Robotics. Decentralized science could benefit greatly from a better system and infrastructure that allows for access to these technologies in open, cheap, and collaborative way.

AI4DeSci (www.AI4DeSci.com) is an ecosystem that is dedicated to continuously building sustainable infrastructure to support AI-empowered open & collaborative decentralized science. The ecosystem mission aims to build infrastructure that accelerate DeSci, to develop use cases of scientific impact, and promote DeSci through actions. The ecosystem aims to reinforce and facilitate both the Ideation and Actions that happen throughout DeSci project lifetime. This will manifest through the **Collaborative Cloud** and **Science Crowdsourcing** main platforms that are built together with community. The collaborative cloud allows scientist, researcher, students, and general public to leverage the decentralized independent cloud infrastructure offered for hosting collaboration. Collaborations that happen can be as simple as writing research paper together, as well as developing analysis protocol, or data processing pipeline. The Science Crowdfunding provide utilities for posting and organizing scientific task distribution campaign.

Collaborative Cloud

The Collaborative Cloud will be focusing on allowing researcher to **Collaborate-as-you-go** with research project that they intend to share and gain support/partnership from the public. Specifically, we are focusing the early effort on facilitating collaboration between scientific use case that would benefit from advance AI algorithm. For instance, there are plenty of biologist, material scientist, and scientist alike that are not natively trained on developing state-of-the-art AI tools or algorithm. At the same time there are great many AI Scientist and AIOps (Ops for IT Operations), that would be very excited to contribute to use cases that are way more meaningful for humanity progress forward, not for simply optimizing profit or creating more clicks. By allowing for collaboration to happens between these two groups of people we can expect to see great and inventive solution using advanced AI pipeline to solve civilization most challenging problem from energy issues, biotech challenges, personalized medicine and more.

The collaboration that happens will have a great degree of openness and inclusion to be decided no other than by the project owner/contributors themselves. No mandates from the "higher up" would decide what degree should the project be proceeding at. We believe decentralized science should at its heart empower those who do science. To enable this from ground up, the infrastructure that serves for this collaboration should also be decentralized. This is necessary to ensure no-one tampered with the precious data and account of this scientific collaboration, nor that someone can simply paywalled or boycott the project by expensive cost of infrastructure. Specifically, the cloud environment in which the discussions, the code being developed, and the data produced throughout should be on decentralized cloud and on decentralized storage. Great platform like Akash Network has enable these forms of decentralized compute, along with Filecoin and Arweave that allows for storing huge data on public domain that are decentralized and at the same time well-protected are indispensable in the success of Decentralized Science moving forward.

That being said, many efforts to integrate AI and Science would require significant amount of high-quality data. These data, if to be generated through the traditional pipeline will cost enormously both in terms of time and capital. Thus, in the next section we delve deeper onto how Crowdsourcing Science will be one possible option to remediate the trade-off between cost and availability of the quality data needed for the endeavor.
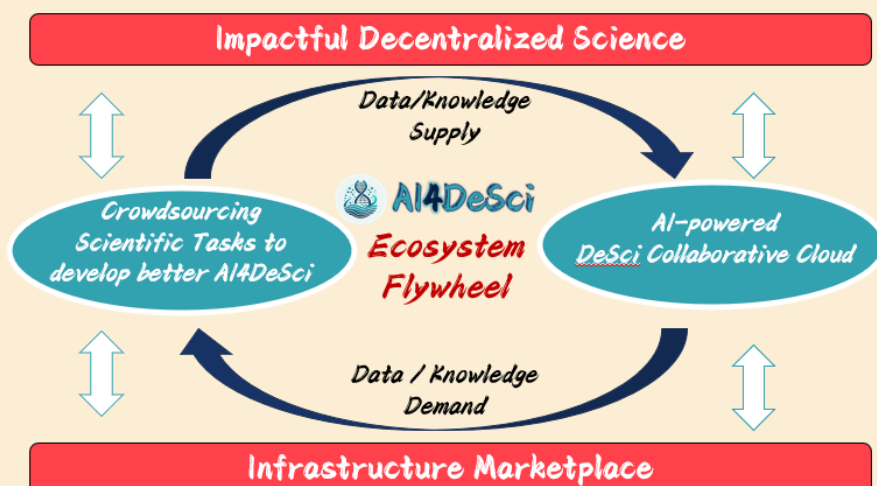
Crowdsourcing Science

Building a useful and meaningful AI Agents or ML model requires a great deal of scientifically sound data. One need a great deal of "Textbooks" and "Syllabus" in order to teach these AI models techniques or knowledge useful for scientific domain application. But like any of our great textbooks and syllabus of education, there are numerous great contributions of individual from across generations needed. Luckily in the digital era these contributions can be gathered in a parallel and widespread manner through the support of internet ecosystem of the 21st century. Moreover, the blockchain technology and tokenomics ecosystem surrounding it has allowed our generation to build various infrastructure and incentive mechanism that can accelerate this progress faster and better than what our previous generation can ever imagine. There's no limit as to what can be crowdsourced. We have seen a great many examples of how this can be done and are useful with project such as Folding@Home, Picbreeder, GalaxyZoo and more that span across discipline and involved not only thousands of people but up to millions over the years. This testify to what extend Crowdsourcing Science could be beneficial.
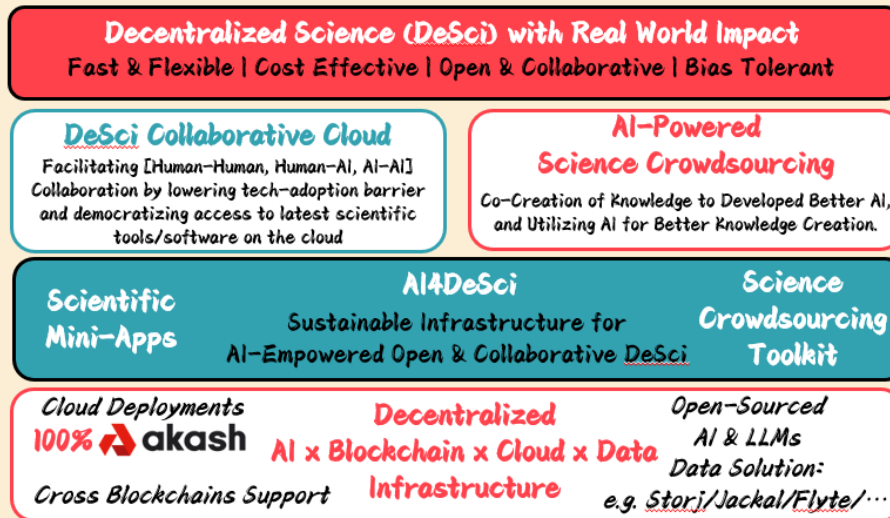
Unfortunately, those great project requires an infrastructure that are built independently and currently a bit outdated, making it hard to simply copy the success onto new projects. Along the process the projects also carried out the mission of public knowledge dissemination that help spawns interest among citizen scientist to also explore then build things and tools that are useful for science. In our platform we decided to first tackle the issue of annotating and labelling scientific data for training models and AI-Agents on scientific matter. Different for the general-purpose labelling and annotation that currently done across the world by third-world cheap labors, or by other bots specialized to do so. Scientific data labelling and annotation would require more rigor. The three projects mentioned earlier, have to certain degree demonstrate this need of scientific rigor. For instance, in the PicBreeder project they tap into the imagination aspect of human that is missing from bots and maybe missing for third-world data labelling army interested only in money. GalaxyZoo educate the public to recognize different pattern of galaxy rotation/expansion before user then can help them categorize the satellite imagery available through the platform. Folding@Home necessitate user to have some basic understanding about protein folding, so that while their computer works on the problem, user can also recognize issue with the process.

Therefore, in our Crowdsourcing Science platform, user can utilize many of the general and customization friendly infrastructure to run their own crowdsourcing science project, to help them with improving data quality as well as empowering the creation of even better state-of-the-art model for scientific applications. More can be done with platform like these and we are just in the beginning of it.
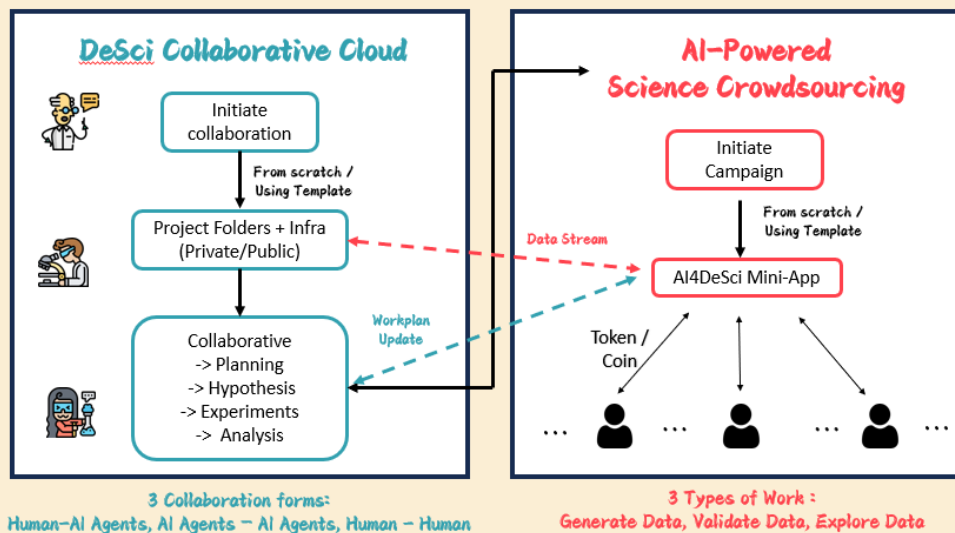
Stay tuned for more.

[This is first working draft for the whitepaper. Tech architecture to support the ecosystem will be shared in the next version update. More references and Case studies would be included to exemplify context] Comment and provide feedback below to improve the whitepaper.

🌐 AI4DeSci *Ecosystem*

Decentralized Science (DeSci) with Real World Impact
Fast & Flexible | Cost Effective | Open & Collaborative | Bias Tolerant

DeSci Collaborative Cloud
Facilitating [Human-Human, Human-AI, AI-AI] Collaboration by lowering tech-adoption barrier and democratizing access to latest scientific tools/software on the cloud

AI-Powered Science Crowdsourcing
Co-Creation of Knowledge to Developed Better AI, and Utilizing AI for Better Knowledge Creation.

Scientific Mini-Apps

AI4DeSci
Sustainable Infrastructure for AI-Empowered Open & Collaborative DeSci

Science Crowdsourcing Toolkit

Cloud Deployments
100% akash

Cross Blockchains Support

Decentralized AI x Blockchain x Cloud x Data Infrastructure

Open-Sourced AI & LLMs
Data Solution:
e.g. Storj/Jackal/Flyte/···

Main Workflow for Core Platforms

DeSci Collaborative Cloud

Initiate collaboration

From scratch / Using Template

Project Folders + Infra (Private/Public)

Collaborative
-> Planning
-> Hypothesis
-> Experiments
-> Analysis

Workplan Update

Data Stream

AI-Powered Science Crowdsourcing

Initiate Campaign

From scratch / Using Template

AI4DeSci Mini-App

Token / Coin

··· ··· ··· ···

3 Collaboration forms:
Human-AI Agents, AI Agents – AI Agents, Human – Human

3 Types of Work :
Generate Data, Validate Data, Explore Data

Roadmap for AI4DeSci @ Akashathon and beyond

Discovered about Akashathon.
Start learning about Akash.Network

Tech Stack module development,
Use Case Building

Live Deployment of resources and MVP on Akash via Console/Cloudmos

Next Milestones
- 10 In-House Use Cases
- 5 Public Dataset & Models published on AI4DeSci
- 3 DeSci Projects running @AI4DeSci

| 14th April | 15th April | 15th – 30th April | 1st -5th May | 10-17th May | Coming Soon |

AI4DeSci Official Birth
www.AI4DeSci.com

Architecture blueprint, Whitepaper & Docs development

App.AI4DeSci.com
Hub.AI4DeSci.com
First public release

Upcoming Core Features & Integrations

Dominikus Brian 钟鸿盛 | domi@dreambrook.tech