

结合 Web3 的公共物品型向量数据库 与 智能合约 Copilot

— 让向量数据库成为一种公共物品

公共使用 & 公共治理

NonceGeekDAO

2023 / 09



Repo: https://github.com/NonceGeek/movespace_db

Demo: <https://ai.movespace.xyz>

我们将向量数据库作为Web3公共物品，允许公共读取权限，并使用智能合约进行写入治理。

问题:

传统向量数据库的解决方案是本地化部署在服务器上，很难让其成为公共物品，让AI-related 的多个应用程序共享。

愿景:

构建基于Web3的向量数据库**范式**，打造Web3*AI的基础设施，让所有数据贡献者因为利他行为而获益。

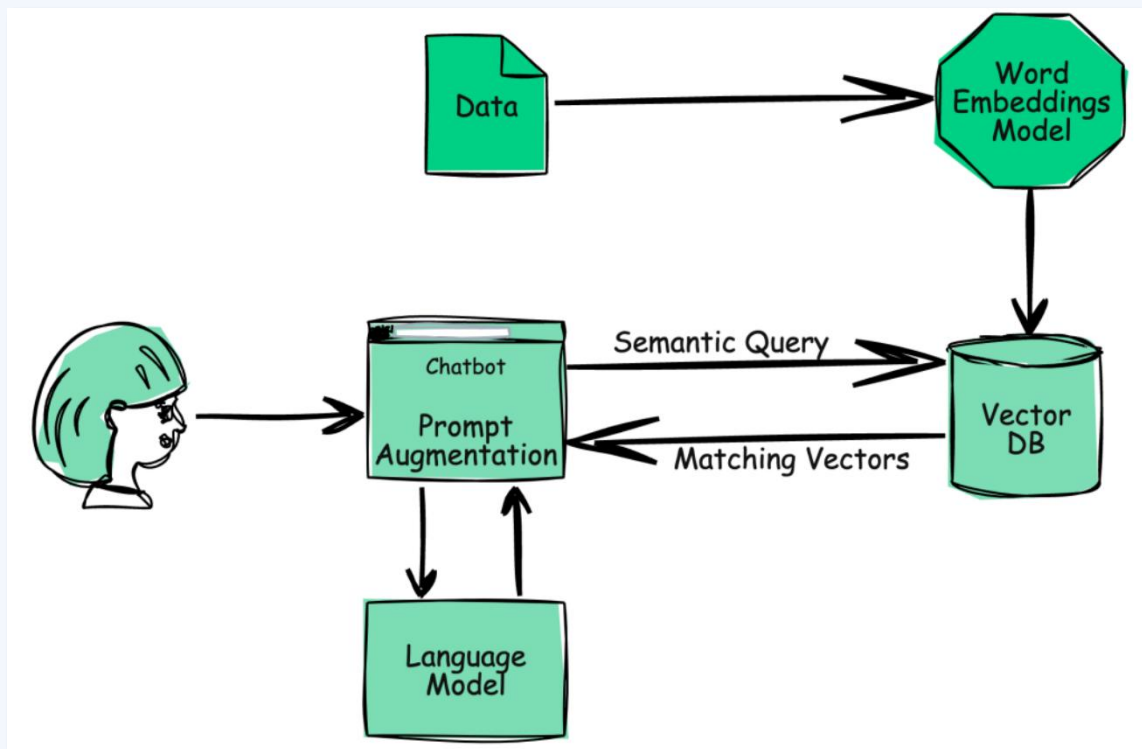


背景梳理



向量数据库的定位与地位

向量数据库技术是人工智能应用中的一项重要技术。使用向量数据库对大型语言模型进行预处理是现有的一个很好的解决方案。

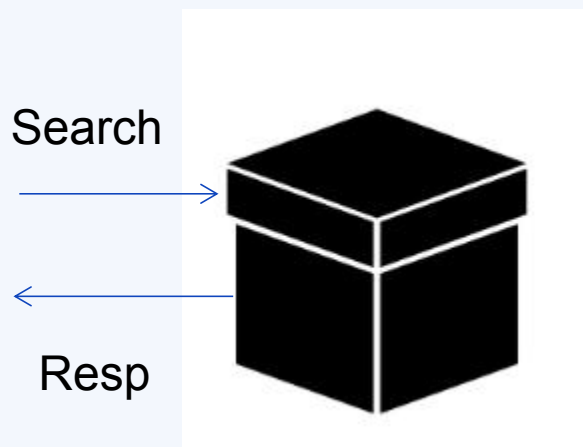




我们为什么需要公共向量数据库？

「社会价值」

面对中心化推荐算法的「难以约束的作恶」，公共向量数据库可以公开透明的「由公众所维护」、「给公众所使用」。



中心化推荐算法，算法不透明





问题现状与解决方案



问题现状与解决方案

问题现状

- 大语言模型缺乏领域知识
- 公共向量数据库的管理尚未成型
- 传统的向量数据库存在单点故障导致数据丢失的可能性

解决方案

- 使用向量数据库存储领域知识（公共或私人托管）
- 使用Web 3*DAO方法进行数据治理
- 使用 Arweave 永存网络进行元数据存储



MoveSpace 架构设计



设计考虑

Large Language Model Selection: GPT-3, GPT-4, the other open source LLM...

Vector Database: pinecone, supabase, weaviate, postgres

Data Write Data Governance Committee: Smart Contract based on Move/Solidity

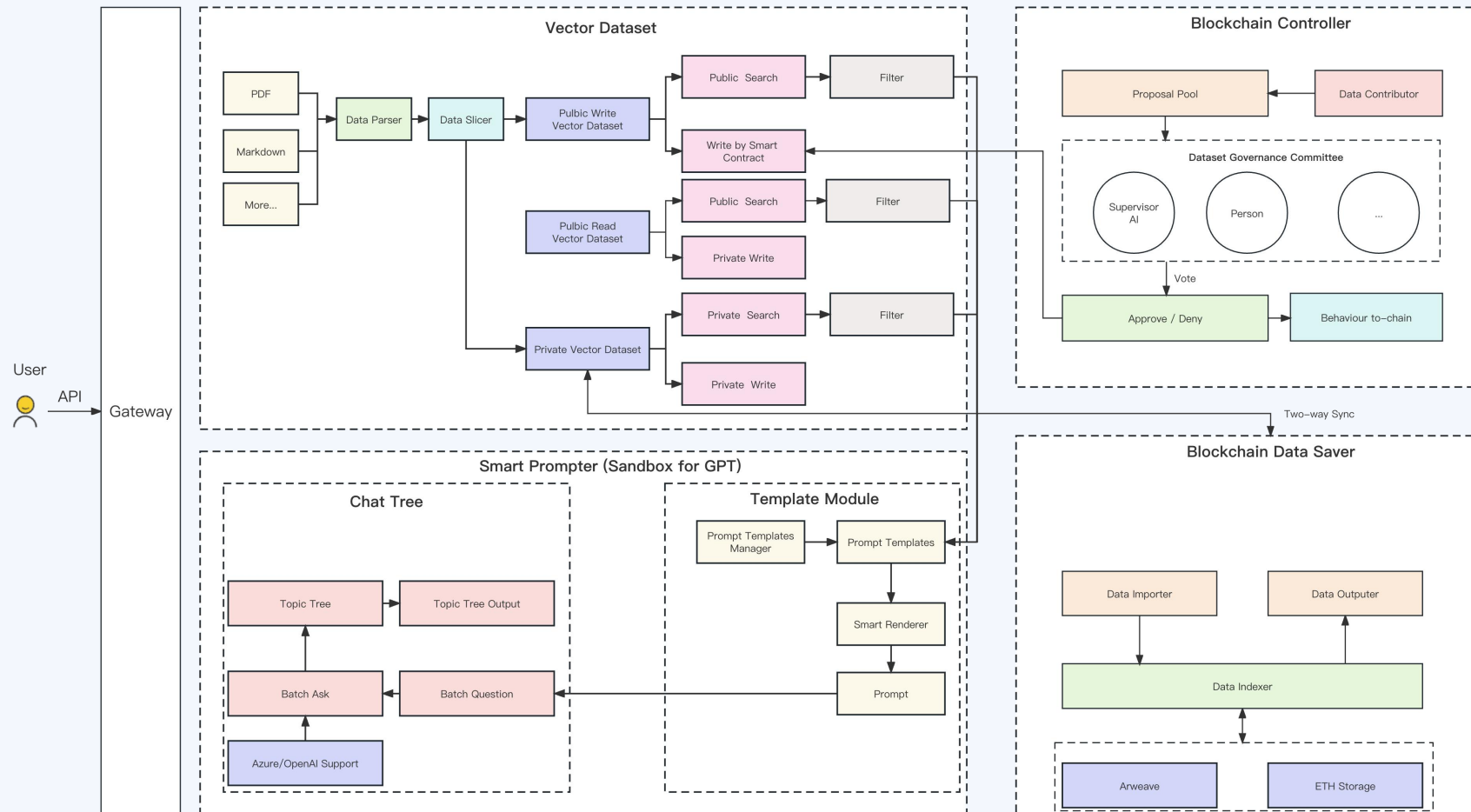
Programming Language: Elixir, Move, Solidity

Data Privacy: Data privacy assurance by AI members in committee



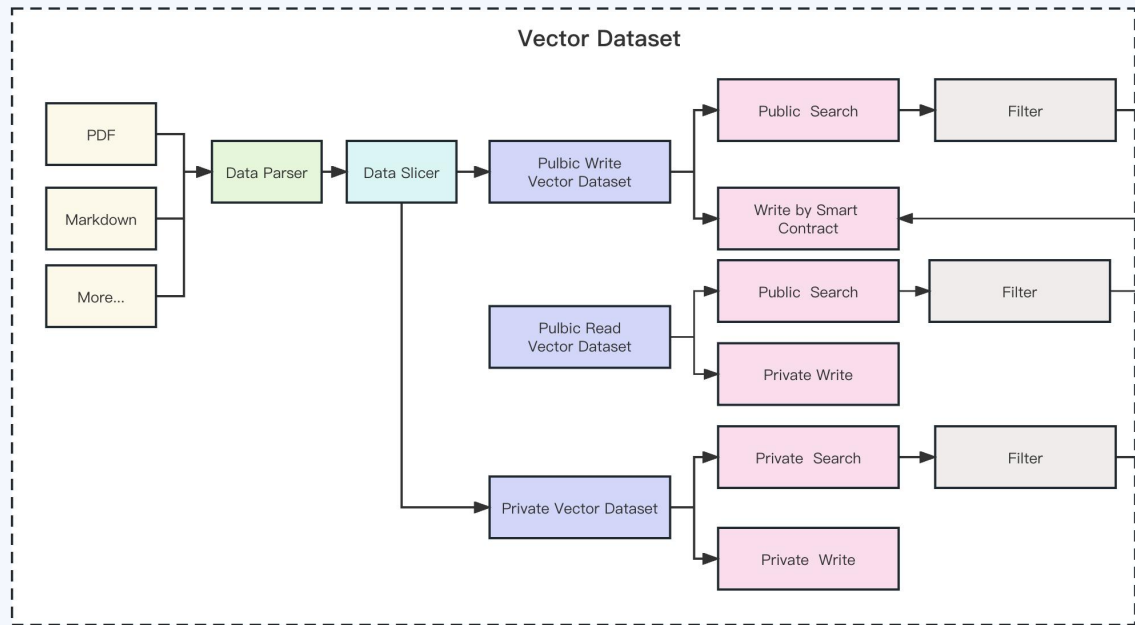
Full Map

四大模块: Vector Dataset, Blockchain Controller, Blockchain Data Saver and Smart Prompter.





Vector Dataset 模块



向量数据库模组提供数据从 PDF、Markdown、PPT 等等多种原始数据源的解析、切片、写入、检索、过滤的完整流程。创建的向量数据库可分为公共写向量数据库、公共读向量数据库、私有向量数据库。

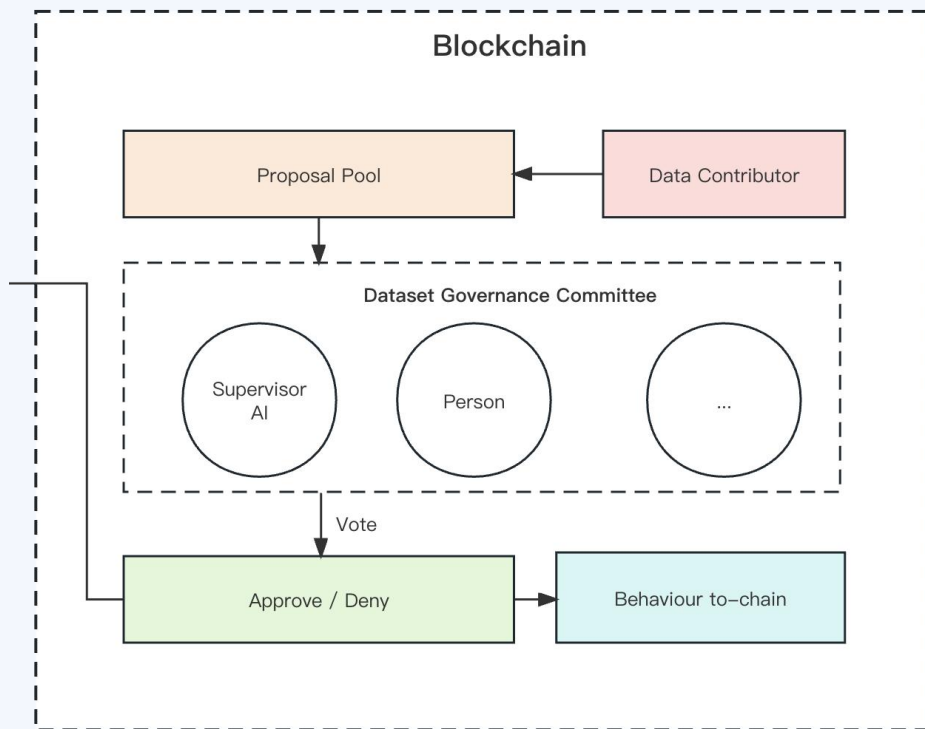
公共写向量数据库完全读开放，并有插件支持，可存储在存储在区块链网络上，从而成为一种稳定的公共服务；在写入方面，公共写向量数据库通过智能合约对写入数据的合规性与合理性进行把控，也即通过区块链模组。公共写向量数据库合于希望多方写入的公共物品类的向量数据库，例如开源代码等。

公共读向量数据库完全读开放；在写入方面则像传统数据库一样进行权限控制。公共读向量数据库适合于由单一机构提供数据的向量数据库，例如统计数据等。

私有向量数据库读写权限均类似于传统数据库。私有向量数据库适合于具有自建向量数据库需求的企业。



Blockchain Controller Module



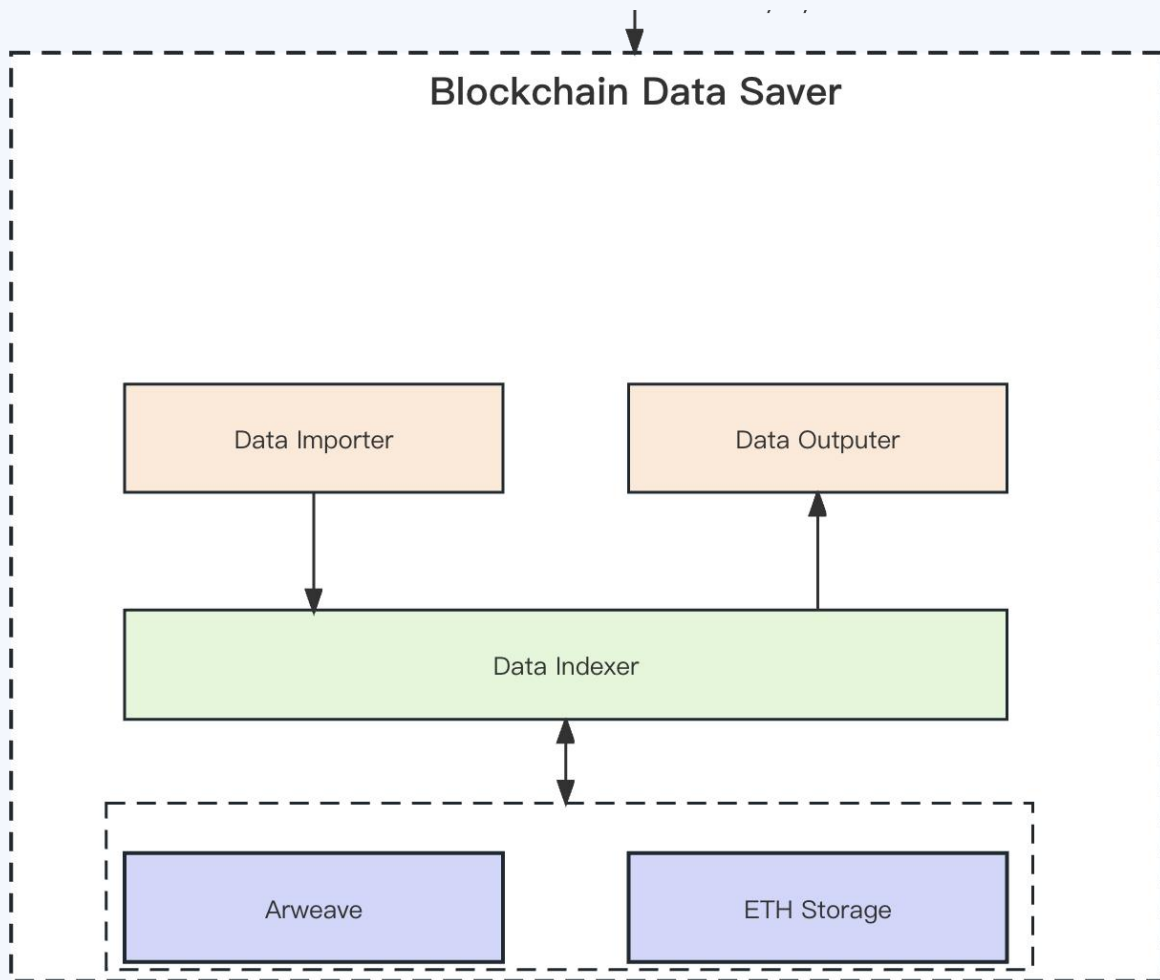
区块链模组提供向量数据库写入权限治理的能力。在贡献者提交申请（Proposal）后，数据库管理委员会（在向量数据库创建之初设置）会对申请进行审核。

数据库管理委员会的成员可以是实体，也可以是 AI。例如，隐私保护 AI、数据合规性检查 AI 等等。

无论最终申请是批准还是否决，提交过程、投票结果和最终审核都会进行上链，从而确保行为在链上的可信记录。



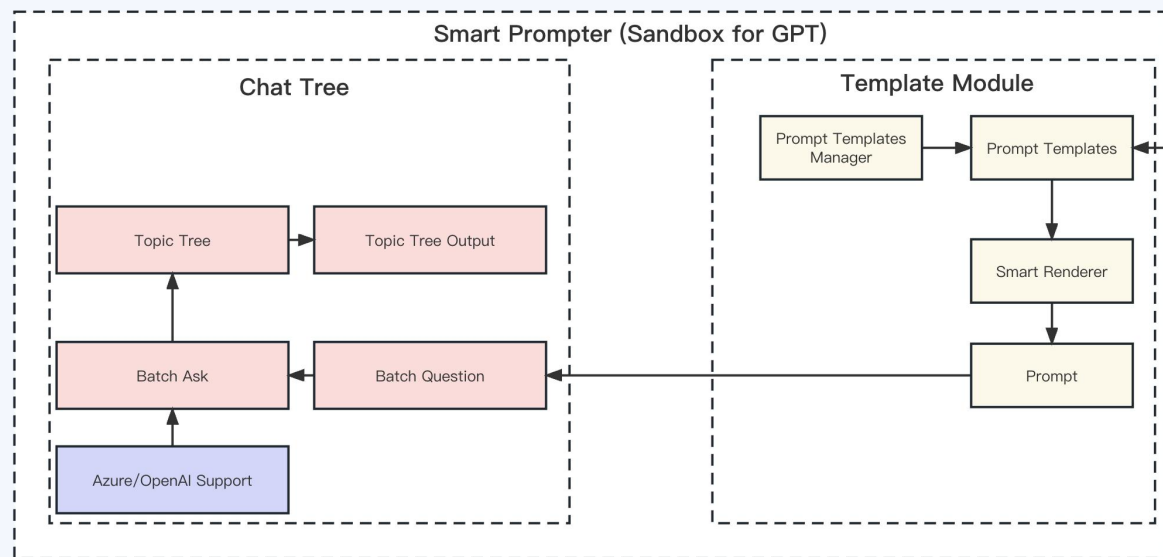
Blockchain Data Saver Module



通过永存型网络保存向量数据库的数据，确保在单点故障的情况下向量数据库依然可以恢复，具备超强稳健性。



Smart Prompter Module



Smart Prompter 模组提供和大语言模型进行对话的能力。在获得向量数据库的最终检索结果后，Smart Prompter 基于检索结果和 Prompt 模板生成 Prompt，生成的 Prompt 可以喂给各类大语言模型。

此外，Smart Prompter 还提供对话树功能，从而可以用逻辑清晰的树状结构和大语言模型进行对话，提高智能程度。



AI Scaffold ——

一个公共向量数据库的例子



基于向量数据库的智能合约生成脚手架

Demo: <https://scaffold.movespace.xyz>

智能生成 ETH 合约或 Move 合约，并给出参考依据源码

Here is an example of a struct called AddressAggregator:

```
struct AddressAggregator {  
    addresses: vector::Vector<Address>,  
    total_addresses: u64,  
}
```

In this struct, AddressAggregator has two fields:

addresses: This field is of type vector::Vector<Address> and stores a collection of Address objects. The vector::Vector is a built-in Move data structure that allows for dynamic sizing and efficient storage of elements.

total_addresses: This field is of type u64 and represents the total number of addresses stored in the addresses vector.

You can use this struct to aggregate a collection of addresses and track the total number of addresses stored.

Search Results in Dataset:

RESULT	METADATA
struct Empty has drop, store {}	%{file_name: "common_account", type: "struct"}
struct MessageChangeEvent has drop, store { from_message: string::String, to_message: string::String, }	%{file_name: "hello_blockchain", type: "struct"}
struct IterableTable<K: copy + store + drop, V: store> has store { inner: TableWithLength<K, IterableValue<K, V>>, head: Option<K>, tail: Option<K>, }	%{file_name: "iterable_table", file_path: "data_structures", type: "struct"}
struct IterableValue<K: copy + store + drop, V: store> has store { val: V, prev: Option<K>, next: Option<K>, }	%{file_name: "iterable_table", file_path: "data_structures", type: "struct"}
struct MessageHolder has key { message: string::String, message_change_events: event::EventHandle<MessageChangeEvent>, }	%{file_name: "hello_blockchain", type: "struct"}

Give me the examples about the Struct. Generate a Struct called AddressAggregator

Get Smart Answer ↩



重要技术特性总结

The Most Important Feature

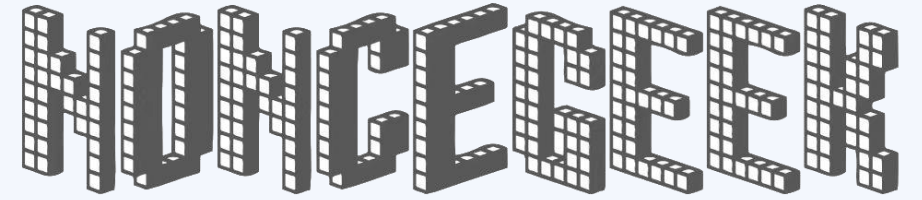
- **公共物品：** 读取权限公开化，为 AI 创新应用提供支撑
- **高稳健性：** 永存网络存储确保向量数据库的高稳健性
- **贡献确权：** 通过治理合约确保数据贡献者被如实记录，成为后续的贡献者激励的基础
- **代码类公共物品：** 首先切入的垂直场景，代码原子化后进入向量数据库，成为开发者的强力辅助工具



团队介绍



NonceGeekDAO



Vision: "Helping Indie Hackers in DAO gain life growth."

Created Time: 2019 -12

Co-Founders: Leeduckgo

Slogan: Cool-oriented Programming.

Github Org: <https://github.com/noncegeek>

DAO: <https://noncegeek.com>

Telegram: @leeduckgo

