Censer Litepaper

目錄

版本

1.0

1 專案概要

Censer 是基於 Polkadot， 採用民主治理的機器學習模型託管服務，相較於傳統由公司帶領產品走向的方式，Censer 將部署權利下放到由開發者、消費者與第三方稽核者組成的 DAO，使「無國界」（borderless）的「多方利害關係人」(Multi Stakeholder)[1]可決定機器學習模型的治理方向以及消費者賠償機制。

Censer 讓每個機器學習模型都是一個 DAO，有專門的存款來承擔再保險的責任，負責管理風險不確定的機器學習模型的部署（和反部署）。所有的部署都要有可驗證聲明智能合約可供第三方稽核者檢驗，檢驗不通過，則執行可驗證聲明智能合約記載的風險控制機制，例如強制停用機器學習模型。

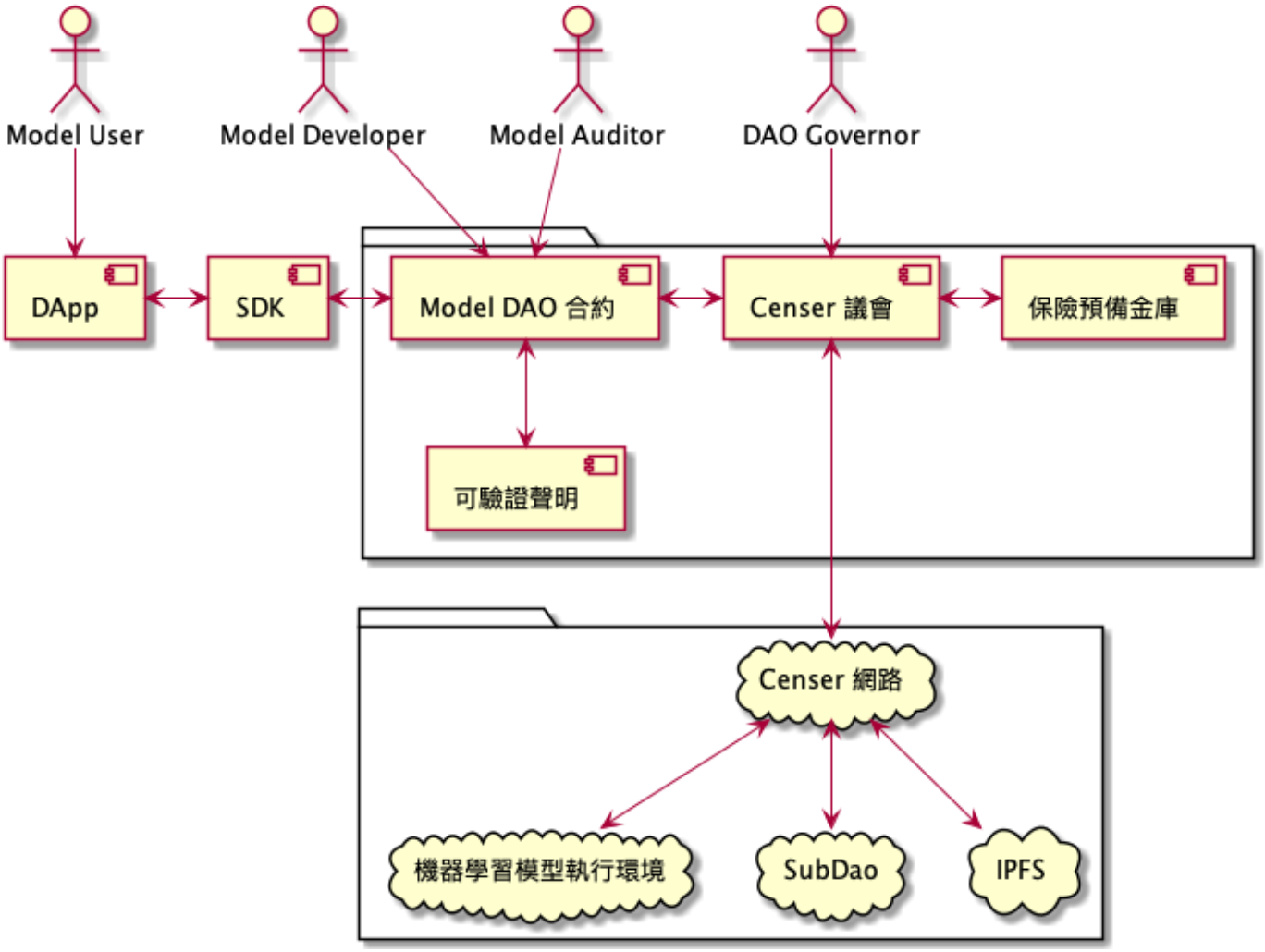
2 動機

本專案的目的在於在法律責任歸屬尚未明時，由機器學習模型的投資者、開發者、使用者提撥一筆資金作為儲備金，藉由智能合約的「無需信任」（trustless）特性，可以先動用儲備金支付損害賠償，使損害受到填補，也分散了機器學習開發者與使用者的責任風險，讓敏感的機器學習模型更容易被市場及監管者接受。

3 架構

Censer 專案將基於 Substrate 3.0、ink!，以及其它 Polkadot 平行鏈來達成目標。Censer 將包含 Censer 節點、Censer 議會、保險預備金庫，以及網頁前端。 DAO 相關的功能透過跨鏈通訊的方式或由 SubDAO 平行鏈實現。

3.1 基礎元件



3.1.1 Censer 節點（Censer Node）

Censer 節點是由 Substrate 3.0 構建的客製化鏈節點，Censer 網路的基礎，包含一般節點的基本功能，同時提供連續部署跟執行機器學習模型的功能。

3.1.2 Censer 議會（Censer Council）

Censer 議會管理整個網路、提供模型開發者建立或註銷 Model DAO。

3.1.3 Model DAO

Model DAO 控制機器學習模型部署，質押有價代幣發行「可驗證聲明」。Model DAO 同時具備基本的投票功能處理消費者受害理賠提案。

3.1.4 可驗證聲明（Verifiable Claims）

可驗證聲明[2]是支付機器學習模型保險金的智能合約，在該合約中會紀錄被保險的機器學習模型，被保險的機器學習模型的執行環境、開發者聲明的特性，以及聲明的特性不符合時需支付的有價代幣。這個可驗證聲明可以被不特定任何人嘗試證偽後取得獎勵，並使機器學習模型返回上一版本或是停止啟用，端看智能合約設定的條件。可驗證聲明將被設計成為 NFT 為鑄造合成資產提供空間。

3.1.5 保險預備金庫（Insurance Reserves Vault）

保險預備金庫是存放機器學習模型保險預備金的智能合約，此合約管理任何有價代幣。

3.1.6 機器學習模型執行環境 (Model Runtime)

機器學習模型執行環境讀取 Model 並做推論，以 Web Service 的方式作為接口。Model Runtime 透過 NixOS 達成可復現，可回滾，在開發端、生產端一致的執行環境。

3.1.7 Web 前端（Web Frontend）

Web 前端提供 Web 介面讓任何人可以跟 Censer 網路互動。前端介面提供基本的部署機器學習模型，質押保險金，轉帳等功能。基於 TypeScript 與 Node.js 開發。

3.1.8 SDK

SDK 讓 Model 開發者能將 DApp 與 Censer 互動。一開始將只支援 TypeScript。

3.2 使用情境

3.2.1 建立 Model DAO

機器學習開發者向議會申請建立 Model DAO。Model DAO 治理模型預設為 MolochDao [3]。

3.2.2 部署機器學習模型

機器學習模型開發者透過 Model DAO 部署機器學習模型、以及發行可驗證聲明合約。

3.2.3 稽核機器學習模型

機器學習模型稽核者透過可驗證聲明合約，輸入可以證偽的「測試資料集」（Testing Dataset)，取得稽核紀錄。

3.2.4 獎勵

機器學習模型稽核者提供稽核紀錄，比對可驗證聲明合約記載的條件取得稽核獎勵。

3.2.5 賠償

機器學習模型使用者提出賠償提案，賠償提案通過 Model DAO 審議後支付賠償金。

3.2.6 下架/回滾機器學習模型

可驗證聲明合約在保險預備金不足時，強制執行是將機器學習模型下架，或是返回到上一版本。

4 Substrate / Polkadot 整合

Censer 會以平行鏈的方式連接上 Polkadot。

4.1 Off-chain Worker

Censer 的節點會在 Off-chain Work 啟用的狀態下編譯。Off-chain Work 用於取得鏈外的資訊，例如機器學習模型與資料集；操作機器學習模型執行環境。

4.2 ink!

前面提到的智能合約將基於 ink! 開發，並運行在客製化的 pallet_contract。

5 治理代幣

Censer 將發行「XI」治理代幣，用於部署模型跟發行可驗證聲明，投票財政國庫提案跟管理網路。Censer 將要求被部署的模型至少需質押 1 XI ，取「一『息』（XI）尚存」之意。所有被質押的「保險預備金」將透過「分散式金融」（DeFi） 放貸或是流動性挖礦孳息，產生的利息 90% 分配至網路財政國庫，用於建設社群，發展更好的生態系，剩餘 10% 則分配至開發團隊。

模型稽核者可透過執行稽核獲得 XI 代幣; 模型使用者可透過持續啟用模型的功能來獲得 XI 代幣; 模型開發者則透過使可驗證聲明持續有效來獲得 XI 代幣。因此部署跟管理保險金賠償的權利將隨著模型上線動態調整，促使「多方利害關係人」利益一致。

6 開發路徑

Phase 1

完成上述提到的最小功能的 PoC 提供少數人測試。 功能包括：

1. 機器學習模型推論
2. 可驗證聲明
3. 保險預備金庫
4. 議會
5. 前端
6. Model DAO

Phase 2

支援更多機器學習模型框架跟可驗證聲明範本。

Phase 3

面向公眾提供服務。

腳註:

[1] 多方利害關係人網路治理模式, https://www.twnic.tw/mps/page5.html.

[2] Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims, http://export.arxiv.org/pdf/2004.07213.

[3] MolochDAO Whitepaper, https://github.com/MolochVentures/Whitepaper.

作者: Hsin-Yi Chen
Email: ossug.hychen@gmail.com
Created: 2021-05-06 Thu 04:01